AFRL-RY-WP-TM-2008-1177

FEATURE-ENHANCED, MODEL-BASED SPARSE APERTURE IMAGING

Dr. Müjdat Çetin, Alan S. Willsky, and Dr. John W. Fisher III

Massachusetts Institute of Technology

MARCH 2008
Final Report

STINFO COPY

AIR FORCE RESEARCH LABORATORY
SENSORS DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7320
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

*//Signature//

CURTIS H. CASTEEL, Jr., 1Lt, USAF
Lead ISR Sensor Signal Engineer
Signature & Modeling Branch
Sensors Directorate

//Signature//

MARTIN E. JUSTICE, Chief
Signature & Modeling Branch
Sensors Directorate

//Signature//

STEVEN P. WEBBER, Lt Col, USAF
Deputy, Sensor ATR Technology Division
Sensors Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YY)* March 2008	2. REPORT TYPE Final	3. DATES COVERED *(From - To)* 01 October 2004 – 29 February 2008

4. TITLE AND SUBTITLE FEATURE-ENHANCED, MODEL-BASED SPARSE APERTURE IMAGING	5a. CONTRACT NUMBER
	5b. GRANT NUMBER FA8650-04-1-1719
	5c. PROGRAM ELEMENT NUMBER 62204F

6. AUTHOR(S) Dr. Müjdat Çetin, Alan S. Willsky, and Dr. John W. Fisher III	5d. PROJECT NUMBER 6095
	5e. TASK NUMBER 19
	5f. WORK UNIT NUMBER 60951900

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Massachusetts Institute of Technology Laboratory for Information and Decision Systems 77 Massachusetts Avenue Cambridge, MA 02139	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory Sensors Directorate Wright-Patterson Air Force Base, OH 45433-7320 Air Force Materiel Command United States Air Force	10. SPONSORING/MONITORING AGENCY ACRONYM(S) AFRL/RYAS
	11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S) AFRL-RY-WP-TM-2008-1177

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
Report contains color. The production of the material in this report was completed under a grant by researchers at Massachusetts Institute of Technology. The material wasn't cleared by the WPAFB Public Affairs Office because it was already in the public domain or is available to the public.

14. ABSTRACT
This report summarizes the research and activities under the Revolutionary Automatic Target Recognition and Sensor Research (RASER) Grant FA8650-04-1-1719 on the topic of Feature-Enhanced, Model-Based Sparse Aperture Imaging. This project has been motivated by a number of emerging military applications where we are faced with sparse apertures. Examples include wide-angle imaging, foliage penetration radar, bistatic imaging, and passive radar imaging. While the possibility of exploiting such rich sensor data presents remarkable opportunities for surveillance, image formation and visualization from sparse aperture data poses significant challenges. The focus of our research effort has been to meet these challenges and develop principled and practical sparse aperture imaging techniques which generate enhanced imagery facilitating visual or automatic interpretation of the underlying scenes. In this report, we provide a picture of the activities and progress that have occurred in this project. In particular, we include both basic factual information on personnel, publications, and interactions, as well as a brief description of our research activities and how they relate to the statement of work included in our proposal. *See Full Abstract on reverse* →

15. SUBJECT TERMS
Synthetic aperture radar, image formation, sparse-aperture imaging, regularization, wide-angle imaging, anisotropy characterization, feature-enhanced imaging, inverse problems, superresolution, anisotropy, sparse signal representation, overcomplete dictionaries, optimization methods, tree searching, Hough transforms, hyper-parameter selection, direction-of-arrival estimation, overcomplete representation, sensor array processing, source localization, sparse representation, coherent imaging, image reconstruction, regularization, computed imaging, half-quadratic regularization, nonquadratic optimization, SAR, wide-angle, sparse measurements, edge-preserving regularization

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT: SAR	18. NUMBER OF PAGES 126	19a. NAME OF RESPONSIBLE PERSON (Monitor) Curtis H. Casteel, Jr.
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER *(Include Area Code)* N/A

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18

14. ABSTRACT (Full)

This report summarizes the research and activities under the Revolutionary Automatic Target Recognition and Sensor Research (RASER) Grant FA8650-04-1-1719 on the topic of Feature-Enhanced, Model-Based Sparse Aperture Imaging. This research has been performed at the Massachusetts Institute of Technology. The primary researchers who have led this research program are Dr. Müjdat Çetin, Prof. Alan Willsky, and Dr. John Fisher.

This project has been motivated by a number of emerging military applications where we are faced with sparse apertures. Examples include wide-angle imaging, foliage penetration radar, bistatic imaging, and passive radar imaging. While the possibility of exploiting such rich sensor data presents remarkable opportunities for surveillance, image formation and visualization from sparse aperture data poses significant challenges. The focus of our research effort has been to meet these challenges and develop principled and practical sparse aperture imaging techniques which generate enhanced imagery facilitating visual or automatic interpretation of the underlying scenes.

In this report, we provide a picture of the activities and progress that have occurred in this project. In particular, we include both basic factual information on personnel, publications, and interactions, as well as a brief description of our research activities and how they relate to the statement of work included in our proposal.

1 INTRODUCTION

This report summarizes the research and activities under the Revolutionary Automatic Target Recognition and Sensor Research (RASER) Grant on the topic of Feature-Enhanced, Model-Based Sparse Aperture Imaging. This project started on October 1, 2004, and ended on February 29, 2008. During he course of this project, we believe that we have accomplished our objectives by making progress in various dimensions:

 (i) We have developed new algorithms in line with the statement of work in our proposal.

 (ii) We have defined student theses topics of direct relevance for this project. Four different students in three different institutions have made significant contributions to the project.

(iii) We have established collaborations and interactions with various research groups, which contributed greatly to the wealth of ideas involved in our work.

(iv) We have made an effort to present our work in various venues to assure its impact. In particular, we have made sure to interact with many colleagues from AFRL, and made them aware of our work.

 (v) Towards the end of this project, we got involved in a new MURI grant funded by AFOSR that is well-aligned with the goals of this research effort.

In this report, we provide brief descriptions of our activities and research, and refer the reader to the publications listed at the end of the report.

1.1 Statement of Work

The main goal of our research effort is to develop principled and practical sparse aperture imaging techniques which generate enhanced imagery facilitating visual or automatic interpretation of the underlying scenes. Our starting point and foundation in this effort is the recently-developed *feature-enhanced imaging* framework. Our proposal included the following statement of work, under full-level of funding:

- **Task 1:** Adapt and demonstrate the effectiveness of feature-enhanced imaging in a variety of 2D/3D sparse aperture imaging scenarios and modalities. Perform quantitative analysis of the formed imagery, and explore the impact of various sensing factors on imaging and illuminate the resulting tradeoffs.

- **Task 2:** Develop techniques for automatic selection of parameters involved in feature-enhanced sparse aperture imaging.

- **Task 3:** Develop image formation strategies that take into account the anisotropic nature of the scatterers.

- **Task 4:** Develop extensions of feature-enhanced imaging to make it robust to errors in sensing model parameters.

- **Task 5:** Explore the enhancement of various kinds of features of importance in sparse aperture imaging applications.

- **Task 6:** Consider extension of feature-enhanced imaging to problems involving non-linear scattering models.

- **Task 7:** Extend feature-enhanced imaging to incorporate data from multiple sensors with imperfect knowledge of the sensor model parameters.

- **Task 8:** Consider visualization strategies that preserve the high-dimensional information inherent in the reconstructed scenes.

- **Task 9:** Develop procedures for decision-oriented sparse aperture imaging, where certain aspects of the image formation algorithm are driven by feedback from the final decision-making objectives.

We have also set priorities in this statement of work, as a function of the available funding level, which we paste below:

- Option 3 [Funding level: $300,000]: With this cost option (**which is the actual level of support we have received in this project**), our research agenda will be as follows. Regarding the work in Task 1, we will mostly focus on 2D radar imaging applications and evaluation of the formed imagery. We will perform the tasks in Task 2 and Task 4. Finally, we would also be interested in carrying out at least some aspects of the research described in Task 3.

- Option 2 [Funding level: $375,000]: With this cost option, our research agenda will include the following items in addition to those described under Option 3. We will consider a variety of 2D imaging modalities in Task 1. We will perform the tasks described in Task 3, as well as those in Task 5.

- Option 1 [Funding level: $450,000]: With this cost option, our research agenda will include the following items in addition to those described under Option 2. For the work in Task 1, we will be able to consider 2.5-3D imaging problems for a variety of modalities. This will benefit from our interaction with AFRL researchers, and hopefully from the use of some of the computational facilities at AFRL. This option will also let us spend some time on the work described in Tasks 6-9 as well.

Given that our funding has stayed at the Option 3 level, we have focused on Tasks 1,2,3,4, and also made some contributions to Task 5. However, thanks to the AFOSR MURI effort we have recently got involved in, we are in the process of formulating research problems to address issues under Task 9 as well.

1.2 Personnel and Data

The primary researchers involved in this research program are Dr. Müjdat Çetin, Prof. Alan Willsky, and Dr. John Fisher. A graduate student, Kush Varshney, has been directly supported by this project. Three students have contributed to this project without receiving direct support. When the project started Müjdat Çetin was a full-time Research Scientist at MIT. In September 2005, he took a faculty position at Sabancı University, İstanbul, Turkey. However, he continued to honor his commitment to this project by: 1) continuing to be affiliated officially with MIT as a Research Affiliate, 2) continuing to be involved in supervising graduate student work on

this project, 3) continuing his collaborative work (e.g. with Ohio State and Boston University) that is relevant for this project, 4) being present at MIT in the summers to conduct research, 5) initiating research at Sabancı University that is related to this effort, without receiving direct support.

Our collaborators, who have contributed to this project without receiving direct support include Prof. Randy Moses (The Ohio State University), Prof. W. Clem Karl (Boston University), Dr. Rajan Bhalla (SAIC), Dr. Thomas Kragh (Lincoln Laboratory), Dr. Eugene Lavely (BAE Systems Advanced Information Technologies), and Prof. Aaron Lanterman (Georgia Tech).

Throughout our work, we have made extensive use of the "Backhoe Data Dome," distributed by AFRL, as part of the VISUAL-D program. We have also used some XPATCH data provided to us by SAIC. We have also obtained the "2D/3D Imaging GOTCHA Challenge Problem" public release data for use in some pieces of our work.

2 RESEARCH ACCOMPLISHMENTS

2.1 Methods for Imaging from Wide-Angle, Sparse-Aperture Data

As mentioned in our statement of work, our starting point for tackling sparse aperture imaging problems has been the feature-enhanced imaging framework. Due to the observation model-based nature of this framework, it has been possible to extend and apply it to various sparse aperture imaging scenarios. One of our major efforts along this line has been our collaborative work with Prof. Randy Moses of the Ohio State University on wide-angle synthetic aperture radar (SAR) imaging.

This work has its center of gravity in **Task 1** and **Task 3**. We have developed a subaperture-based imaging approach for wide-angle SAR where in each subaperture we perform feature-enhanced imaging and then put together all these subaperture images into a composite image for visualization. This approach addresses some of the limitations of conventional, polar format imaging in the context of wide-angle data. Polar format algorithm works on the assumption that all scatterers in the scene persist through all observation angles, which does not hold in wide angular apertures (e.g. 100 degrees). As a result, scatterers that persist only in a small angular range are suppressed. Yet, these scatterers can represent important features of the scene. Our approach alleviates this problem and preserves scatterers with short persistence. In addition, due to its model-based nature, our approach provides much better robustness in the case of aperture omissions (in the frequency band or in the angle band). Since we form subaperture images using feature-enhanced imaging, we also improve the spatial resolvability of scatterers. Finally, our composite images carry not only reflectivity information, but also information on the direction of the maximum scattering response for each scatterer. This provides an additional feature for tasks such as automatic target recognition (ATR). We have presented our results on wide-angle SAR imaging from partial-aperture data with frequency-band omissions at the Algorithms for SAR Imagery Conference, part of the SPIE Defense and Security Symposium [1]. We are in the process of preparing and submitting a journal paper describing this work [2]. We are also happy to observe that a number of papers presented at the Algorithms for SAR Imagery Conference in recent years were using or building upon our work, which is a good indication that our research has started to create an impact on the research community.

Another line of work on which we have started to interact with Randy Moses involves feature-enhanced interferometric SAR imaging from data with frequency-band omissions. Prof. Moses and his students have developed an IFSAR imaging technique built upon our feature-enhanced

imaging framework. We are discussing certain ways to extend that work. Prof. Moses and his team have also started processing the GOTCHA data, and in that effort utilized our feature-enhanced imaging algorithm as well. We are also interacting on new ways to process that data.

In addition to wide-angle SAR, we have explored other sparse-aperture imaging scenarios as well. One example is passive radar imaging. In collaboration with Prof. Aaron Lanterman from Georgia Tech, we have developed a region-enhanced, sparse-aperture passive radar imaging technique, and have demonstrated its advantages over conventional imaging. Our work on sparse aperture passive radar imaging has been published in IEE Proceedings Radar, Sonar & Navigation, Special Issue on Passive Radar Systems [3].

2.2   Methods for Joint Imaging and Anisotropy Characterization

Kush Varshney, the only student receiving direct support from this grant, completed his Master's thesis on joint image formation and anisotropy characterization in wide-angle SAR [4]. We are not attaching the thesis to this report due to its large size, however the thesis can be accessed through the following URL:

`http://www.mit.edu/~krv/pubs/krvarshney_sm.pdf`

This work mainly focuses on **Task 1**, **Task 3**, **Task 5**, and has some connections to **Task 9**. The main idea is to perform joint anisotropy characterization and imaging (reflectivity estimation), by posing both problems as sparse signal representation problems using overcomplete dictionaries. Whereas a conventional radar image produces a complex-valued scalar reflectivity for each scatterer, our approach acknowledges that scattering varies with angle, and produces a scattering function for each scatterer. From this information, one can extract features such as scattering direction and angular scattering extent for each scatterer in the scene. This not only makes the imaging (reflectivity estimation) process in the presence of anisotropic scattering more accurate, but it also produces features that are not present in conventional images and that can be useful for automatic target recognition (ATR). One general principle in previous work on the anisotropy problem has been to divide the full wide-angle aperture into smaller subapertures and form a sequence of subaperture images with inherently reduced cross-range resolution for use in further processing. Another general principle has been to develop parametric models for angle-dependent scattering behavior. The proposed methodology does not suffer a reduction in resolution because the entire available aperture is used and is more flexible than parametric models. The proposed framework solves for multiple spatial locations jointly, ameliorating the ill-effects of close proximity neighboring scatterers. A graph-structured interpretation leading towards novel approximate algorithms to solve the inverse problem is developed. These algorithms, having reduced memory requirements, may well find application in a wide variety of sparse signal representation settings beyond the specific problem of anisotropy in SAR. The first conference paper on this work was presented at the Algorithms for SAR Imagery Conference in April 2006. The results show great promise in characterizing complicated anisotropic scattering behaviors likely to be encountered in wide-angle imaging applications.

On top of the basic work included in that paper, we made some further progress. In particular we achieved two extensions of the basic framework in [5], and published that work in [6] in June 2006. The first of these extensions involves migratory scattering centers. Certain scattering mechanisms, such as tophats and cylinders, appear to migrate or move in their spatial location as a function of aspect angle with wide-angle apertures. This type of scattering, which has not been given much heed in past work, is well-incorporated into our overcomplete dictionary

formulation. In the first part of [6], we present an extension of our overcomplete dictionary for characterizing anisotropy to account for migratory scattering. The second extension is based on the interesting relationship between anisotropy and physical extent in the spatial domain. Scattering response over only a very small range of aspect angles, known as glint or flash, arises from long, flat plates, and the thinner the anisotropic response, the longer the spatial extent of the plate. The aspect angle of the glint is also the orientation of the object in space. In the second part of [6], utilizing Hough transform properties, we introduce new regularization terms to favor solutions that concentrate the representation of glint anisotropy across a spatially distributed area into a single scatterer. Through such extensions to the sparsifying regularization cost function, certain object-level preferences are essentially encoded within the image formation process. This is a principled attempt towards the objective of decision-directed imaging, exploiting high-level information in front-end signal processing. We were invited to present our work at a special session on Radar and Sensor Signal Processing at the 2007 IEEE Conference on Signal Processing and Communications Applications [7]. We have submitted a journal paper based on this work to IEEE Transactions on Signal Processing [8], which was accepted. We expect this paper to appear some time in 2008.

During Müjdat Çetin's stay at MIT during the summers, we got involved in one other piece of work on anisotropy characterization in collaboration with Prof. Clem Karl and his student Ivana Stojanovic. This work was motivated in part by the following observation. Let us consider the other two major pieces of work we have discussed so far, in particular the subaperture-based wide-angle imaging work of Section 2.1, and the joint imaging and anisotropy characterization work described above. Let us evaluate how these two approaches constrain the anisotropy structure across angle. The subaperture-based approach puts no constraints on that. Each subaperture is processed independently, and then, for each scatterer, we stick together reflectivity estimates across angle to get a rough estimate of the angular scattering function. On the other hand the work described above in this section puts a very strong structural constraint on the angular anisotropy. In particular, it only allows angular responses that can be expressed in terms of a pre-selected dictionary for basic scattering mechanisms. The question then was whether we could do something in between, that is, we would like to put a constraint an angular scattering but do not want that to be too strong. We came up with a formulation in which one can perform joint imaging and anisotropy characterization, where angular scattering functions are constrained to be piecewise smooth. This appears to be a reasonable constraint. We have developed algorithms for implementing this idea, and have obtained interesting results on the backhoe data set. We presented this work at the Algorithms for SAR Imagery Conference in 2008 [18].

2.3 Methods for Hyperparameter Choice in Regularization-based Imaging

Dr. Çetin has started to supervise a graduate student, Özge Batu, at Sabancı University, whose Master's thesis topic involves the problem of automatic hyper-parameter choice for regularization-based sparse aperture imaging problems. This work provides contributions to **Task 2**. Regularization-based sparse aperture imaging techniques combine mathematical models of the data collection process with contextual information about the scene to be imaged. When such pieces of information are combined in the right manner, these techniques provide robust and feature-enhanced reconstructions, providing significant improvements over conventional imaging approaches. Yet, this requires manually selecting some hyper-parameters that establish the balance between different pieces of information. For widespread and seamless use of such imaging algorithms in

practice, techniques for automatic hyper-parameter choice are needed. This research is aimed at developing such techniques. We have looked into existing automatic parameter choice techniques applied in different fields, with a particular focus on non-quadratic problems. Two particular techniques, Stein's unbiased risk estimator (SURE), and generalized cross validation (GCV), appear to be promising in terms of their potential performance for sparse aperture imaging applications. We have adapted these technique to sparse aperture radar imaging problems.

Both SURE and GCV involve computational difficulties when considered in the framework of non-quadratic regularization-based imaging. First of all, they require the computations involving large scale matrix multiplications and inversions which are not practical at all. In addition, both methods require the solution of an optimization problem over the hyperparameter. We have developed a number of numerical techniques to address these issues. We have applied our parameter choice techniques on the backhoe data, as well as on various synthetic data collection scenarios. We observe that these techniques can provide reasonable, but slightly underregularized solutions. This has been a very important first step towards fully automatic processing in feature-enhanced sparse aperture SAR imaging. We presented this work at the Algorithms for SAR Imagery Conference in 2008 [9]. A preliminary version of this work has also been presented at a conference in 2007 [10].

We have also developed a new algorithm for solving sparse signal representation problems [11]. This algorithm might be extended to address certain aspects of the problem of automatic selection of parameters involved in feature-enhanced sparse aperture imaging (**Task 2**) as well. A paper that describes our work is [11].

2.4   Methods for Joint Imaging and Model Error Correction

Dr. Çetin has started to supervise another graduate student, Özben Önhon, at Sabancı University, whose Ph.D. thesis topic is focused on the problem of sensing model errors in sparse aperture imaging scenarios. This work mainly provides contributions to **Task 4**. Model-based sparse aperture imaging requires the use of a mathematical model of the data collection process for effective scene reconstruction. Yet, in many scenarios, there are uncertainties in the observation model, e.g., due to imperfect knowledge of the position of the sensing platform. Such model errors lead to various artifacts in the reconstructed images, which could have adverse effects, e.g. on the performance of the ATR system that utilizes these images. This research aims to develop imaging algorithms that exhibit robustness to such errors. The modality of particular interest is SAR. For SAR, existing autofocus-based techniques for dealing with model errors are not satisfactory in a sparse aperture imaging context. These techniques rely heavily on conventional image formation, and view the best model parameter estimate as the one that improves the conventional image in a particular fashion. Yet, in sparse aperture imaging contexts, conventional images are often not of acceptable quality, even if there are no model errors. Consequently, there is a need to consider the imaging and model correction problems jointly, rather than as consecutive steps. We have formulated the problem of joint sparse aperture imaging and model error correction as a joint optimization problem. We have obtained some preliminary results on synthetic data, which demonstrate the potential of this approach in correcting model errors. We are in the process of writing a paper describing this work [12].

2.5 Connections to Follow-up Work on ATR

We got involved in an AFOSR MURI titled "Integrated Fusion, Performance Prediction, and Sensor Management for Automatic Target Exploitation (ATE)." That effort has clear ties to the work supported by this grant. In particular, innovative front-end processing involving advanced imaging techniques, forms an important component of that project. As a result, we believe that the MURI effort will constitute an important follow-up activity after this particular research effort, and will utilize the progress we have made. We are in the process of formulating some research problems involving the interaction of various components of an ATE system. This includes, for example, the concept of "decision-directed imaging" which is the topic of **Task 9**.

2.6 Other Parts of our Sparse-Aperture Imaging Work, including Application to Sensing Modalities other than SAR

While our primary focus in developing sparse-aperture imaging and signal processing techniques in this project has been SAR, our mathematical framework and algorithms have the potential to be useful in other scenrarios and application domains as well. We were involved in a number of such activities ourselves, and we provide very brief information on some of them here.

We have developed a superresolution technique for source and target localization with acoustic (possibly sparse) sensor arrays. This work has been published in IEEE Transactions on Signal Processing [13].

We have extended our work on SAR to other coherent imaging modalities. This work has been published in Optical Engineering [14].

We have applied the techniques we have developed to sparse aperture ultrasound imaging for non-destructive evaluation. A paper describing our results has been published at the IEEE International Conference on Acoustics, Speech, and Signal Processing [15].

3 INTERACTIONS WITH OTHER RESEARCHERS

A number of our collaborators have contributes to this project without receiving direct support. These include Prof. Randy Moses (The Ohio State University), Prof. W. Clem Karl (Boston University), Dr. Rajan Bhalla (SAIC), Dr. Thomas Kragh (Lincoln Laboratory), Dr. Eugene Lavely (BAE Systems Advanced Information Technologies), and Prof. Aaron Lanterman (Georgia Tech). We have already described our collaborative work with Prof. Moses, Prof. Karl, and Prof. Lanterman, which has involved numerous visits and meetings. Here we briefly mention the remaining interactions.

We have had a beneficial interaction with Dr. Rajan Bhalla from SAIC. Dr. Bhalla's past work on electromagnetic scattering and anisotropy characterization has both commonalities and complementary aspects with our perspective. So his perspective on our work has been very valuable for us. Dr. Bhalla provided us some XPATCH data, which we used effectively in our work on joint imaging and anisotropy characterization discussed in Section 2.2. In addition, our discussions with Dr. Bhalla have motivated our work on migratory scattering centers [6]. Overall, this interaction has provided benefits for **Task 1**, **Task 3** and **Task 5**.

We have interacted with Dr. Thomas Kragh from MIT's Lincoln Laboratory. Dr. Kragh has previously used some of our algorithms on a number of radar imaging problems, and has been interested in our work supported by this grant. He has performed some analysis of one of our

imaging algorithms, and published that work at the IEEE International Conference in Image Processing in 2006.

We have been involved in a synergistic activity with Dr. Eugene Lavely from BAE Systems Advanced Information Technologies (formerly Alphatech, Inc.), through a subcontract from an AFRL SBIR grant, where the goal is to use the feature-enhanced imaging ideas developed in this project, as a foundation for feature-based tracking and ATR based on multi-sensor data.

4   PROFESSIONAL ACTIVITIES AND IMPACT

Dr. Çetin has organized a special session at the 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing, on the topic of "Advances in Sparse Signal Representation," which is a topic that forms the basis of the sparse-aperture imaging algorithms we have developed under this grant. This conference was held on March 19-23, 2005, in Philadelphia. This session brought together prominent experts working on the theory and applications of this topic, and was attended by around 150 people.

Dr. Çetin has delivered a one-hour talk at the Workshop on Imaging from Wave Propagation at the Institute for Mathematics and its Applications (IMA) of the University of Minnesota. Dr. Çetin's talk included some of the work supported by this grant. This workshop was part of the IMA Thematic Year on Imaging, and was held on October 17-21, 2005. Many colleagues from AFRL were also attendees at this workshop, with whom we had fruitful discussions about our work and AFRL's interests.

Dr. Çetin has served as a panelist at the Algorithms for SAR Imagery Conference in April 2007. The panel topic was the GOTCHA Challenge Problem of imaging in an urban sensing environment, in which the sensor collects data about the scene over extended periods of time.

Dr. Çetin was an invited speaker to present the work on joint image formation and anisotropy characterization at a special session on Radar and Sensor Signal Processing at the IEEE Conference on Signal Processing and Communications Applications in June 2007 [7].

The 2008 Algorithms for SAR Imagery Conference, which is part of the SPIE Defense and Security Symposium, contained a special session titled "Sparse Recognition for Imaging." We are happy to observe that our work performed under this project has provided inspiration for the topic of this special session. Furthermore, most of the papers presented in this session contained direct references to our work. Finally, there was a panel discussion on using sparsity for radar, which we feel has been inspired in part by various pieces of our sparsity-driven imaging work for SAR. Overall, we are happy to see that our work has had some impact and our colleagues are using various parts of our ideas and algorithms for making further progress on radar imaging in general, and sparse-aperture imaging in particular.

5   PUBLICATIONS

The following is a list of recent papers, theses, and other publications connected with the research conducted under this grant.

References

[1] M. Çetin and R. L. Moses, "SAR imaging from partial-aperture data with frequency-band omissions," in *Algorithms for Synthetic Aperture Radar Imagery XII*, E. G. Zelnio and F. D. Garber, Eds., Orlando, FL, USA, Mar.-Apr. 2005, Proc. SPIE.

[2] M. Çetin and R. L. Moses, "Point-enhanced subaperture-based imaging for wide-angle SAR," *IEEE Trans. Aerosp. Electron. Syst.*, in preparation for submission.

[3] M. Çetin and A. D. Lanterman, "Region-enhanced passive radar imaging," *IEE Proceedings Radar, Sonar & Navigation, Special issue on Passive Radar Systems*, vol. 152, no. 3, pp. 185–194, June 2005.

[4] K. Varshney, "Joint image formation and anisotropy characterization in wide-angle SAR," M.S. thesis, MIT, Cambridge, MA, May 2006.

[5] K.R. Varshney, M. Çetin John W Fisher III, and A. S. Willsky, "Joint image formation and anisotropy characterization in wide-angle SAR," in *Algorithms for Synthetic Aperture Radar Imagery XIII*, E. G. Zelnio and F. D. Garber, Eds., Orlando, FL, USA, Apr. 2006, Proc. SPIE.

[6] K.R. Varshney, M. Çetin John W Fisher III, and A. S. Willsky, "Wide-angle SAR image formation with migratory scattering centers and regularization in Hough space," in *Adaptive Sensor Array Processing Workshop*, Lexington, MA, USA, June 2006.

[7] K.R. Varshney, M. Çetin John W Fisher III, and A. S. Willsky, "A sparse signal representation-based approach to image formation and anisotropy determination in wide-angle radar," in *IEEE Conference on Signal Processing and Communications Applications*, Eskişehir, Turkey, June 2007.

[8] K.R. Varshney, M. Çetin, John W Fisher III, and A. S. Willsky, "Sparse signal representation in structured overcomplete dictionaries with application to synthetic aperture radar," *IEEE Trans. Signal Processing*, to appear.

[9] Ö. Batu and M. Çetin, "Hyper-parameter selection in non-quadratic regularization-based radar image formation," in *Algorithms for Synthetic Aperture Radar Imagery XV*, E. G. Zelnio and F. D. Garber, Eds., Orlando, FL, USA, Mar. 2008, Proc. SPIE.

[10] Ö. Batu and M. Çetin, "Non-quadratic regularization based image deblurring: Automatic parameter selection and feature based evaluation," in *IEEE Conference on Signal Processing and Communications Applications*, Eskişehir, Turkey, June 2007.

[11] D. M. Malioutov, M. Çetin, and A. S. Willsky, "Homotopy continuation for sparse signal representation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, Pennsylvania, Mar. 2005, vol. 5, pp. 733–736.

[12] Ö. Önhon and M. Çetin, "An optimization-based approach to joint advanced imaging and model error correction in SAR," in preparation for submission.

[13] D. M. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[14] M. Çetin, W. C. Karl, and A. S. Willsky, "A feature-preserving regularization method for complex-valued inverse problems with application to coherent imaging," *Optical Engineering*, Jan. 2006.

[15] M. Çetin, E. Bossy, R. Cleveland, and W. C. Karl, "Sparsity-driven sparse-aperture ultrasound imaging," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, May 2006.

[16] Ö. Batu and M. Çetin, "Hyper-parameter selection in advanced synthetic aperture radar imaging algorithms," in *IEEE Conference on Signal Processing and Communications Applications*, Aydın, Turkey, April 2008.

[17] M. Çetin and R. L. Moses, "Synthetic aperture radar imaging from wide-angle data with frequency-band omissions," in *IEEE Conference on Signal Processing and Communications Applications*, Antalya, Turkey, Apr. 2006.

[18] Ivana Stojanovic, M. Çetin, and W. C. Karl, "Joint space aspect reconstruction of wide-angle SAR exploiting sparsity," in *Algorithms for Synthetic Aperture Radar Imagery XV*, E. G. Zelnio and F. D. Garber, Eds., Orlando, FL, USA, Mar. 2008, Proc. SPIE.

[19] K. Varshney, "Wide-angle synthetic aperture radar," 6.961 Introduction to Research in Electrical Engineering and Computer Science Project Report, Massachusetts Institute of Technology, Cambridge, Massachusetts, Dec. 2004.

SAR Imaging from Partial-Aperture Data with Frequency-Band Omissions

Müjdat Çetin[a] and Randolph L. Moses[b]

[a]Laboratory for Information and Decision Systems, Massachusetts Institute of Technology,
77 Massachusetts Ave., Cambridge, MA 02139, USA

[b]The Ohio State University, Department of Electrical and Computer Engineering
2015 Neil Avenue, Columbus, OH 43210, USA

ABSTRACT

We consider the problem of wide-angle SAR imaging from data with arbitrary frequency-band omissions. We propose an approach that involves composite image formation through combination of subaperture images, as well as point-enhanced, superresolution image reconstruction. This framework provides a number of desirable features including preservation of anisotropic scatterers that do not persist over the full wide-angle aperture; robustness to bandwidth limitations and frequency-band omissions; as well as a characterization of the aspect dependence of scatterers. We present experimental results based on the Air Force Research Laboratory (AFRL) "Backhoe Data Dome," demonstrating the effectiveness of the proposed approach.

Keywords: synthetic aperture radar, wide-angle imaging, sparse-aperture imaging, feature-enhanced imaging, inverse problems, superresolution

1. INTRODUCTION

Traditional image formation techniques for synthetic aperture radar (SAR) rely on data on a *narrow-angle*, *filled* aperture. In particular, it is customary to assume that the phase history data lie in an (almost rectangular) annular region in the 2-D spatial frequency domain, establishing a filled synthetic aperture in both the angle (azimuth) and the frequency (range) direction. This is based on the fact that many traditional systems integrate over relatively small angles (typically on the order of a few degrees) and transmit over an uninterrupted portion of the frequency spectrum. However, there are a number of emerging applications where neither of these assumptions holds. One such application is monostatic wide-angle imaging, which may be used to obtain ultra-high resolution at relatively high operating frequencies, or to compensate for the reduced resolution in relatively low frequencies. The data in wide-angle sensing usually lie in a narrow arc in the spatial frequency domain, which constitutes a sparse aperture since the data support fills only a small portion of the circumscribing rectangle. A number of recent technology advancements enable consideration of wide-angle imaging. First, advancements in GPS and INS systems permit collection of coherent data across longer times and flight paths. Second, unmanned air vehicle (UAV) technology and collaboration among UAVs provide a number of wide-angle imaging possibilities. UAVs can, in many applications, fly closer to the scene of interest, and thus can traverse a wider-angle aperture in a given amount of time compared to a platform with a greater standoff distance. A second application of interest is foliage penetration (FOPEN) radar, which operates at the VHF/UHF bands. At these relatively low frequencies, it is likely that we will not be able to use an uninterrupted frequency band, due to the existence of other in-band radiators and FCC licenses. As a result, the data will contain frequency-band omissions resulting in a non-traditional, sparse (or at least not filled) aperture. More broadly, partial aperture data involving omissions in the frequency band may be encountered in higher frequencies as well, due to a number of reasons including jamming and data dropouts. A third application involves bistatic and multistatic imaging. One scenario is a bistatic/multistatic radar operation, in which a distant standoff platform acts as the transmitter and one or more

---

UAVs act as (closer-in) receivers. UAVs working in tandem can collect angular subapertures which can then be combined into a wider aperture which potentially involves omissions in the frequency and/or angle bands.

When traditional image formation techniques are applied to wide-angle data with frequency-band omissions, they often yield unsatisfactory results, making the resulting images difficult to interpret and of limited value for further processing. This is due to a number of reasons. First, the point spread function (PSF) of an isotropic point scatterer coherently imaged over a wide-angle aperture is more irregular than the more customary sinc-like PSFs encountered in traditional SAR imaging, leading to sidelobes that might interfere with other scatterers in the scene. Second, when there are frequency-band omissions, the PSFs resulting from conventional imaging become even more irregular, causing yet more pronounced artifacts in the reconstructed images. Furthermore, different types of band omissions lead to different kinds of artifacts, making it a very challenging task to adapt to and interpret the formed imagery. Third, in a wide-angle imaging scenario, the isotropic point scattering assumption employed by conventional imaging does not usually hold, as many scatterers do not persist over such wide apertures and exhibit some aspect dependence. In such a scenario, conventional imaging can lead to inaccuracies in relative reflectivities of scatterers with different levels of anisotropy. Furthermore, such processing only produces a reflectivity estimate of each scatterer but does not characterize its aspect dependence. Yet, such aspect dependence (if accurately extracted) can itself be an important feature for scene interpretation, e.g. for target recognition.

Motivated by these observations, we explore new image formation strategies for wide-angle data with frequency-band omissions. In particular, we consider the combination of two ideas based on our previous work: composite wide-angle image formation based on subaperture images,[1] and model-based, point-enhanced superresolution imaging.[2] Composite image formation aims to address the issue of limited scattering persistence in wide-angle imaging. The idea is to form subaperture images from narrower-angle subsets of the data, and then form a composite image through a nonlinear combination of these subaperture images. When conventional Fourier transform-based imaging is used to form the subaperture images, composite imaging still suffers from artifacts due to the irregular PSFs, especially in cases involving low-bandwidth data or frequency-band omissions. To address these issues, we propose using point-enhanced imaging[2] to form the subaperture images. This technique uses an explicit model of the observation process (hence incorporates information about the structure of the partial aperture), and as a result, is more robust to data limitations. Furthermore this framework also allows the incorporation of prior information about the underlying scene, which can lead to superresolution. Given such point-enhanced subaperture images, we again form a composite image. This imaging strategy produces not only a reflectivity estimate for each spatial location, but also some information on aspect dependence. We present experimental results based on the Air Force Research Laboratory (AFRL) "Backhoe Data Dome,"[3] demonstrating the effectiveness of the proposed approach.

2. WIDE-ANGLE IMAGING WITH FREQUENCY-BAND OMISSIONS

Let us consider a wide-angle imaging scenario with a center frequency of 10 GHz, an angular aperture of 110°, and a bandwidth of 500 MHz. Fig. 1 shows the magnitude image and the frequency support of the simulated Hamming-windowed data from an isotropic point scatterer in such a scenario. We now use this example to discuss image formation strategies from such data as well as partial aperture data with frequency-band omissions.

2.1. Coherent Integration with an Isotropic Scattering Assumption

The conventional processing we consider here interpolates the phase history data lying on a narrow arc to a Cartesian grid, performs zero-padding to fill the circumscribing rectangle (see Fig. 1), and then takes an inverse 2-D Fourier transform to reconstruct the image. Such processing of the data shown in Fig. 1 leads to the PSF in Fig. 2(a). The curved data support leads to this shape of the PSF which is quite different from the sinc-like PSFs of traditional narrow-angle SAR. This PSF is indicative of the types of artifacts that are likely to appear in conventional images of isotropic scatterers from wide-angle data. Here we have assumed that we have the entire 500 MHz band of data. Now let us consider the case where we have omissions in the frequency band. In particular, let us consider the two masks in Fig. 3, indicating two patterns of band omissions leading to 70% and 30% of the data being available, respectively. In Fig. 2(b) and 2(c) we show the PSFs that result from conventional imaging in the case of such partial aperture data with frequency-band omissions. These PSFs
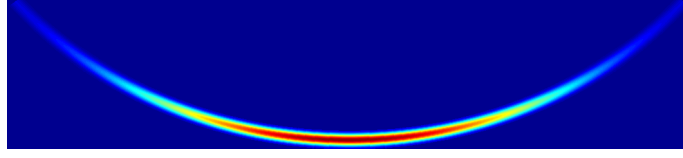
Figure 1. Magnitude image and frequency support of Hamming-windowed data from an isotropic point scatterer over a 110° aperture. Center frequency is 10 GHz and bandwidth is 500 MHz.
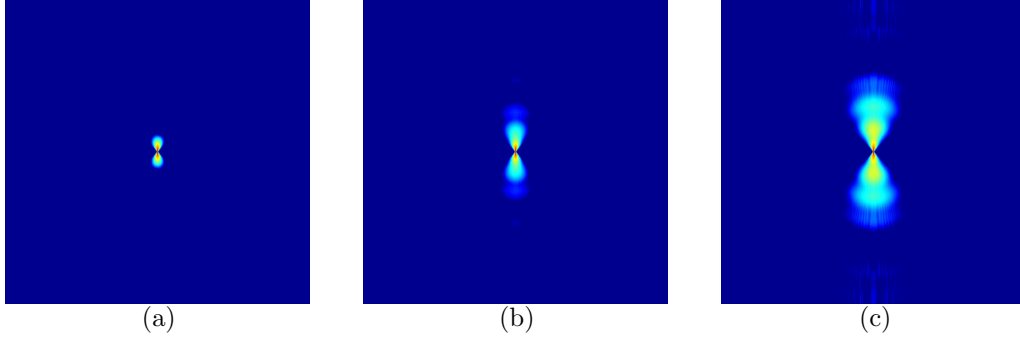

| (a) | (b) | (c) |

Figure 2. Conventional images of a point scatterer based on the data shown in Fig. 1. The images show a region of 10×10 meters. Vertical and horizontal directions in the images correspond to range and cross-range respectively. The images are in logarithmic scale and show the top 40dB of the responses. (a) Full frequency band available. (b) 70% of the frequency band available (based on the mask in Fig. 3(a)). (c) 30% of the frequency band available (based on the mask in Fig. 3(b)).

exhibit significant, wide lobes, suggesting that conventional imaging will cause severe artifacts in these scenarios. In these illustrative examples, we have considered an isotropic point scatterer. Of course, another problem with conventional imaging is that most scatterers will not persist over such wide-angle apertures, and the isotropic scattering assumption will fail. This is an issue we address in the next section.

2.2. Composite Image Formation

In order to accommodate the aspect dependence of the scatterers, we have considered a composite image formation strategy in Ref. 1, which we summarize next. The idea is to use a bank of K matched filters, each characterized by a center response azimuth and a response width and shape.[4, 5] Each of these matched filter outputs is an image conventionally reconstructed from a subaperture of the full azimuth aperture. The underlying assumption is that it is reasonable to assume isotropic scattering within the angular extent of these subapertures. Given the subaperture images $\hat{\mathbf{f}}^k$ for all subapertures $k \in \{1, ..., K\}$, the composite image $\hat{\mathbf{f}}$ is formed as follows:

$$\hat{\mathbf{f}}_{ij} = \arg \max_k \hat{\mathbf{f}}_{ij}^k \tag{1}$$

where $\hat{\mathbf{f}}_{ij}^k$ and $\hat{\mathbf{f}}_{ij}$ denote the (i,j)-th pixel of the k-th subaperture image and of the composite image, respectively. Thus, the composite image has the interpretation of a Generalized Likelihood Ratio Test (GLRT) statistic for scattering responses with known response shape[4, 5] but with unknown peak response angle. We note that in addition to the reflectivity estimates, there is more information available at the output of this process, namely for each pixel we know the index k of the corresponding subaperture image at which the maximum occurs. This provides some characterization of the aspect dependence of the scatterers, which may be useful for aiding object visualization or for use in an automatic target recognition algorithm.
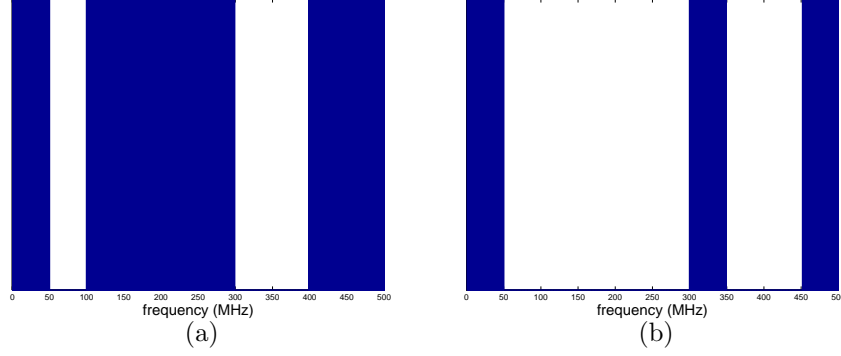
Figure 3. Two patterns of frequency-band omissions, dark regions indicating bands where data are available and light regions indicating missing bands. The masks in (a) and (b) lead to 70% and 30% of the data from the full band being available, respectively.
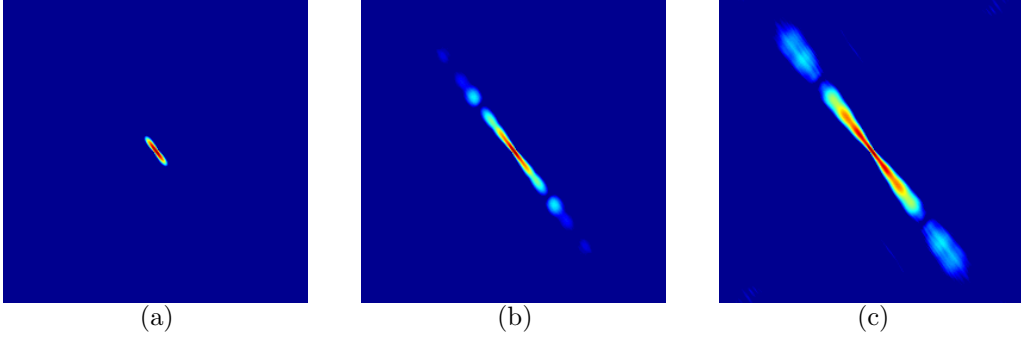

Figure 4. Conventionally reconstructed images of a point scatterer based on data from a $20°$ subaperture centered at $45°$. The images show a region of 10×10 meters. (a) Full frequency band available. (b) 70% of the frequency band available. (c) 30% of the frequency band available.

For illustration, let us view the PSFs corresponding to a single subaperture image that would then be used in composite image formation. In particular, let us consider a subaperture of the data shown in Fig. 1, which is centered at $45°$ and which has a width of $20°$. The PSF for the case of no frequency-band omissions is shown in Fig. 4(a), which is essentially a sinc-like response wider in the range direction than in the cross-range direction. The PSFs for the frequency-band omissions corresponding to the two patterns in Fig. 3 are shown in Fig. 4(b) and 4(c). We note that frequency-band omissions cause significant widening of the PSFs, implying that if conventionally formed subaperture images are used in composite image formation, the final image will suffer from significant artifacts. In the next section, we consider an alternative strategy to address this issue.

2.3. Model-based, Point-enhanced Composite Image Formation

For subaperture image formation, we consider an approach based on the feature-enhanced image formation framework of Ref. 2. In particular, in this paper we focus on resolving and enhancing spatially-localized features, and consider the point-enhanced imaging idea of Ref. 2. This imaging technique can use data in the phase history, the range profile, or the spatial domain. Here we consider the version where we use the conventional image as the input data, hence the technique works as a deconvolution method. In particular, let $\mathbf{y}^k$ be the conventionally reconstructed k-th subaperture image, and let $\mathbf{H}^k$ be a matrix each row of which contains a spatially shifted version of the corresponding PSF (stacked as a row vector). Then point-enhanced subaperture imaging is achieved by solving the following optimization problem:

$$\hat{\mathbf{f}}^k = \arg \min_{\mathbf{f}} \left\{ \|\mathbf{y}^k - \mathbf{H}^k \mathbf{f}\|_2^2 + \lambda \|\mathbf{f}\|_1 \right\} \tag{2}$$

14

where λ is a scalar parameter. The first term in the objective function of Eqn. (2) is a data fidelity term, incorporating the mathematical model of the observation process $\mathbf{H}^k$ into imaging. The second term enforces sparsity of the reconstructed image, which can lead to superresolution in the case of scenes containing a relatively small number of spatially-localized scatterers. The optimization problem in Eqn. (2) can be solved by using efficient iterative algorithms. We note that this expression is written in matrix-vector form for convenience, however in practice we avoid explicitly forming the large matrices $\mathbf{H}^k$ (hence we reduce the memory requirements), by noting that the matrix vector products can be carried out by convolutional operations. Given such point-enhanced subaperture images, we again form composite images as described in Section 2.2, with the only change of replacing the conventional subaperture images with the point-enhanced ones. Note that this procedure again produces more than just an image of reflectivities, since we also obtain a characterization of the aspect dependence of each scatterer.

3. EXPERIMENTAL RESULTS

We present 2D image reconstruction experiments based on the AFRL "Backhoe Data Dome, Version 1.0," which consists of simulated wideband (7-13 GHz), full polarization, complex backscatter data from a backhoe vehicle in free space.[3] The backscatter data are available over a full upper 2π steradian viewing hemisphere. In our experiments, we use VV polarization data, centered at 10 GHz, and with an azimuthal span of $110°$ (centered at $45°$). We consider four different bandwidths: 500 MHz, 1 GHz, 2 GHz, and 4 GHz. For each of these four bandwidths, we consider both the case of full-bandwidth data, and the case of frequency-band omissions where 70% or 30% of the spectral data within that bandwidth are available. For frequency-band omissions, we use the two masks in Fig. 3 with appropriate scaling to the corresponding bandwidth. For composite imaging, we use 19 subapertures, with azimuth centers at $0°$, $5°$, ..., $90°$, each with an azimuthal width of $20°$. The response shape for each subaperture is chosen to be a Hamming window.

3.1. Linear Aperture

First we consider data that would correspond to a linear flight path of the radar platform. In particular, we use azimuth and elevation pairs that simulate such a linear aperture, with a peak elevation angle (at azimuth center) of $30°$. In Fig. 5 we show images of the backhoe reconstructed from such data with various bandwidths. The composite images in Fig. 5(b) appear to provide larger response amplitudes for narrow-aperture scattering centers as compared to the conventional images in Fig. 5(a). This is because the conventional coherent integration process averages all scatterers (including those with a narrow-angle persistence) over the entire wide-angle azimuthal aperture. We note that these two types of images exhibit similar resolution properties and mainlobe structure for the scatterers. As bandwidth is reduced, some features of the backhoe appear to be lost in the images in Fig. 5(a) and 5(b). In contrast, the corresponding composite, point-enhanced images in Fig. 5(c) appear to preserve and exhibit some of the features present in higher-bandwidth images. We choose the hyperparameter λ in Eqn. (2) by visual assessment of the formed imagery. Automatic hyperparameter choice is a topic of our current research. Next we consider frequency-band omissions. Fig. 6 contains results for the case where 70% of the band is available. We observe that conventional and composite images suffer from sidelobe artifacts, especially in the low-bandwidth cases. On the other hand, composite, point-enhanced images in Fig. 6(c) do not suffer from significant degradations as compared to the full-band versions in Fig. 5(c), exhibiting robustness to frequency-band omissions. Finally, in Fig. 7 we present results for the case where we have only 30% of the frequency band available. All imaging methods exhibit noticeable artifacts in this case, although composite, point-enhanced imaging is still able to localize significant scatterers and features of the backhoe.

3.2. Fixed Elevation

We now consider a different aperture, involving a fixed elevation of $0°$, and present the results of an experimental analysis analogous to the one in the previous section. The results shown in Figs. 8-10 lead to similar observations.
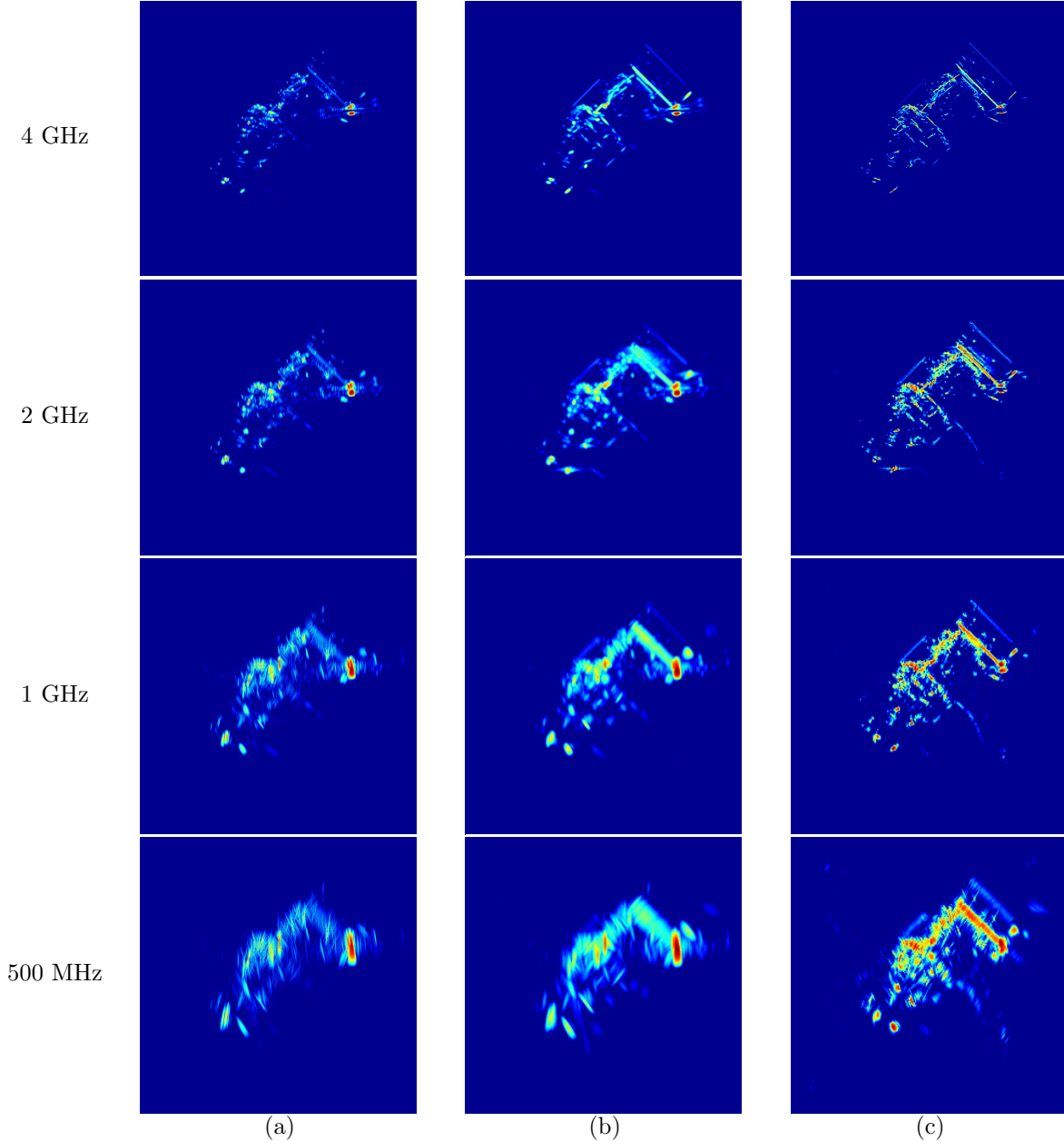
Figure 5. SAR images of the backhoe using a linear aperture and bandwidths of 4 GHz, 2 GHz, 1 GHz, and 500 MHz. (a) Conventional imaging. (b) Composite imaging. (c) Composite, point-enhanced imaging.

3.3. Visualization of Aspect Dependent Scattering

In the composite and composite, point-enhanced reconstruction results in the previous sections, we have only shown the reflectivities at each spatial location. However, as we pointed out in Section 2, we also have the knowledge of which subaperture has led to the maximum reflectivity for each spatial location. This in turn provides some information on the aspect dependence of each scatterer, namely the aspect providing the strongest return from that scatterer. Here we present one way of visualizing that information by encoding the maximum-response aspect through color. In particular, we color-code each pixel by one of 19 colors, corresponding to which of the 19 subapertures identified a maximum. We encode the peak amplitude in the brightness of the pixel. The
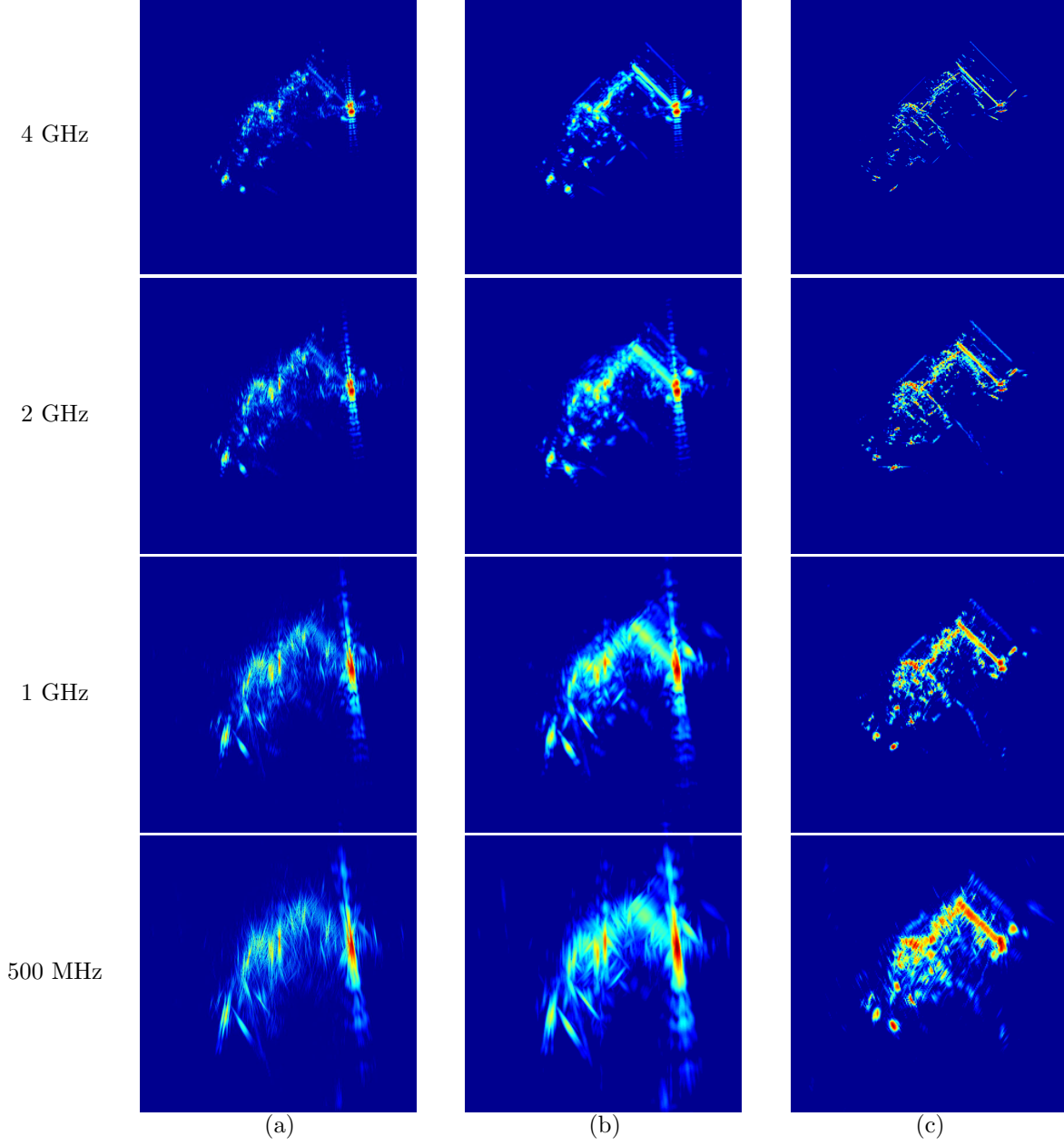
Figure 6. SAR images of the backhoe with frequency-band omissions (70% of the full-band data available) using a linear aperture and bandwidths of 4 GHz, 2 GHz, 1 GHz, and 500 MHz. (a) Conventional imaging. (b) Composite imaging. (c) Composite, point-enhanced imaging.

result is a color image, where red pixels denote maximum response at 0 degrees, green pixels at 45 degrees, and blue pixels at 90 degrees, with colors of intermediate hues representing the aspects in between, resulting in 19 colors each corresponding to a particular aspect. In Fig. 11, we show such color-coded versions of the composite and composite, point-enhanced reconstructions of Fig. 6. These images, especially the point-enhanced ones, suggest that the aspect dependence information extracted in this manner can be informative, and may potentially be useful for scene interpretation, e.g. for target classification.
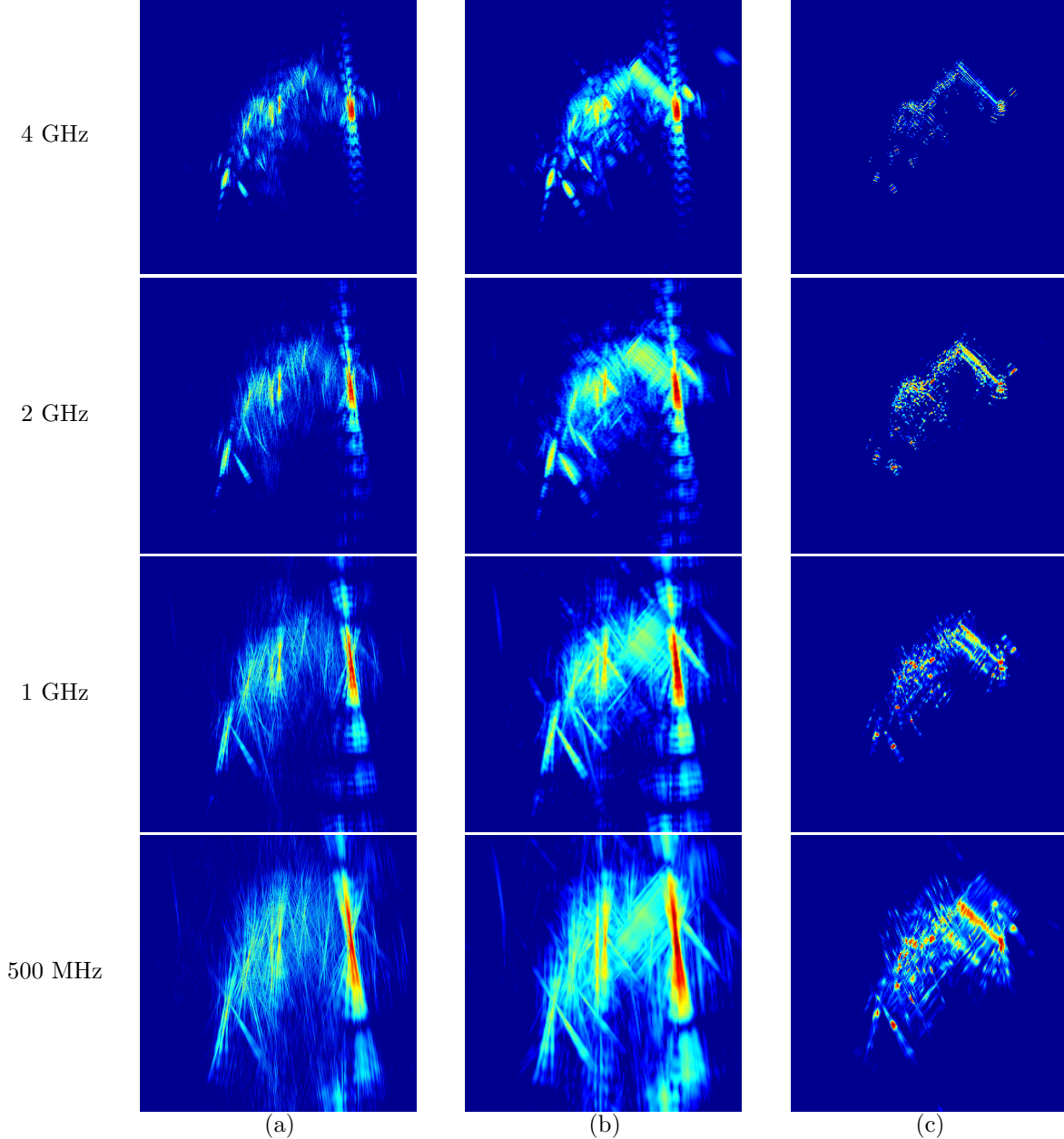
Figure 7. SAR images of the backhoe with frequency-band omissions (30% of the full-band data available) using a linear aperture and bandwidths of 4 GHz, 2 GHz, 1 GHz, and 500 MHz. (a) Conventional imaging. (b) Composite imaging. (c) Composite, point-enhanced imaging.

4. CONCLUSION

We have considered the problem of wide-angle SAR imaging from partial-aperture data with frequency-band omissions. We have proposed an approach that uses model-based, point-enhanced image reconstruction for narrow-angle subapertures, and then performs a nonlinear combination of the subaperture images to form a final wide-angle composite image. We have demonstrated that images formed in this manner exhibit robustness to bandwidth limitations as well as to frequency-band omissions. In addition, this approach yields a partial characterization of aspect dependence, which we have considered visualizing through a color-coding scheme.

Figure 8. SAR images of the backhoe using a 0°-elevation aperture and bandwidths of 4 GHz, 2 GHz, 1 GHz, and 500 MHz. (a) Conventional imaging. (b) Composite imaging. (c) Composite, point-enhanced imaging.

Although we have considered only structured frequency-band omissions in this paper, the approach can also be applied to the case of unstructured omissions, as in random data dropouts. Similarly, these ideas can also be useful for the case of angle-band omissions. One important extension of this work could consider more precise characterization of angular anisotropy, by estimating the persistence level of each scatterer (which was assumed to be equal to the subaperture extent in this paper).

Figure 9. SAR images of the backhoe with frequency-band omissions (70% of the full-band data available) using a 0°-elevation aperture and bandwidths of 4 GHz, 2 GHz, 1 GHz, and 500 MHz. (a) Conventional imaging. (b) Composite imaging. (c) Composite, point-enhanced imaging.

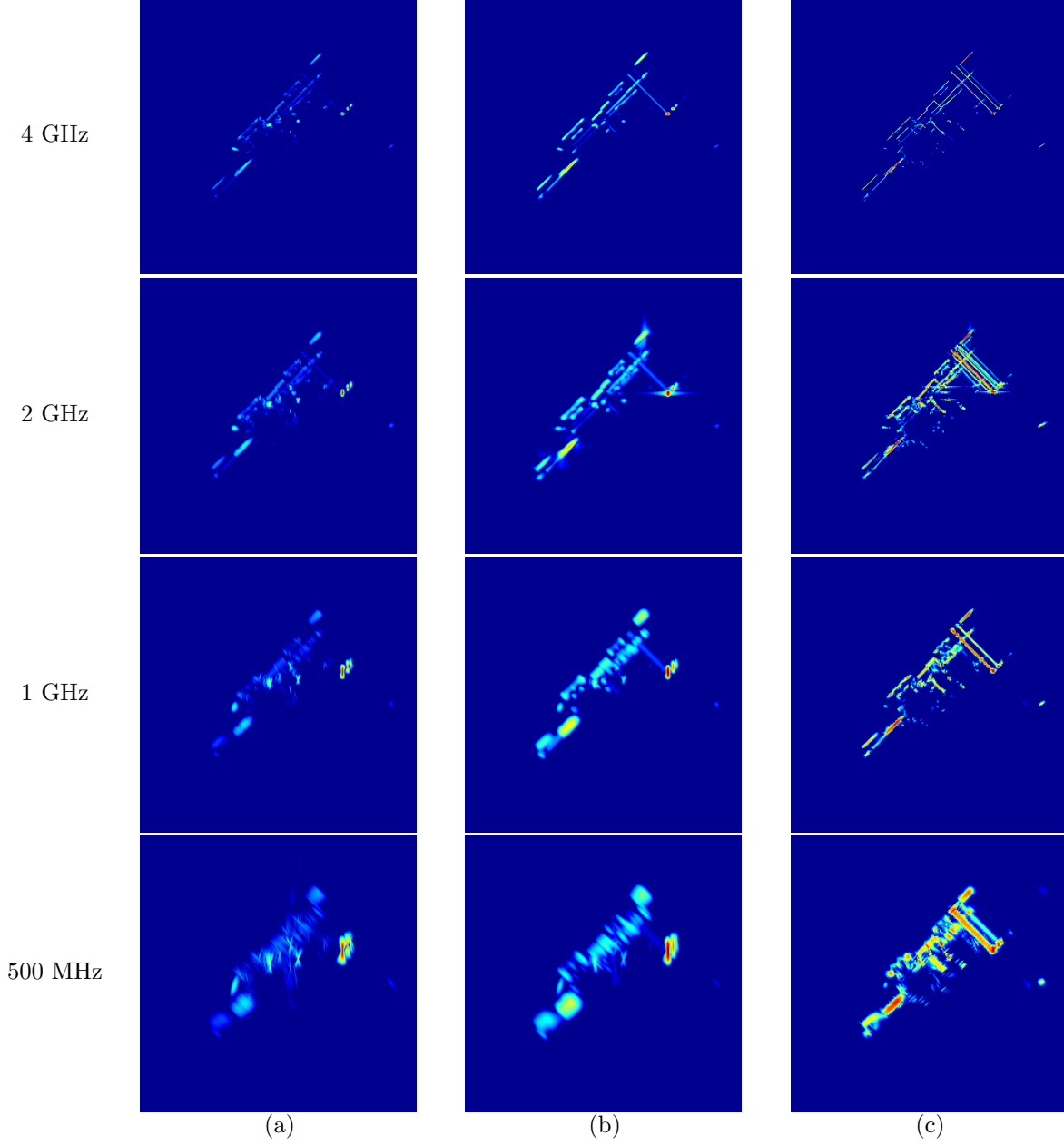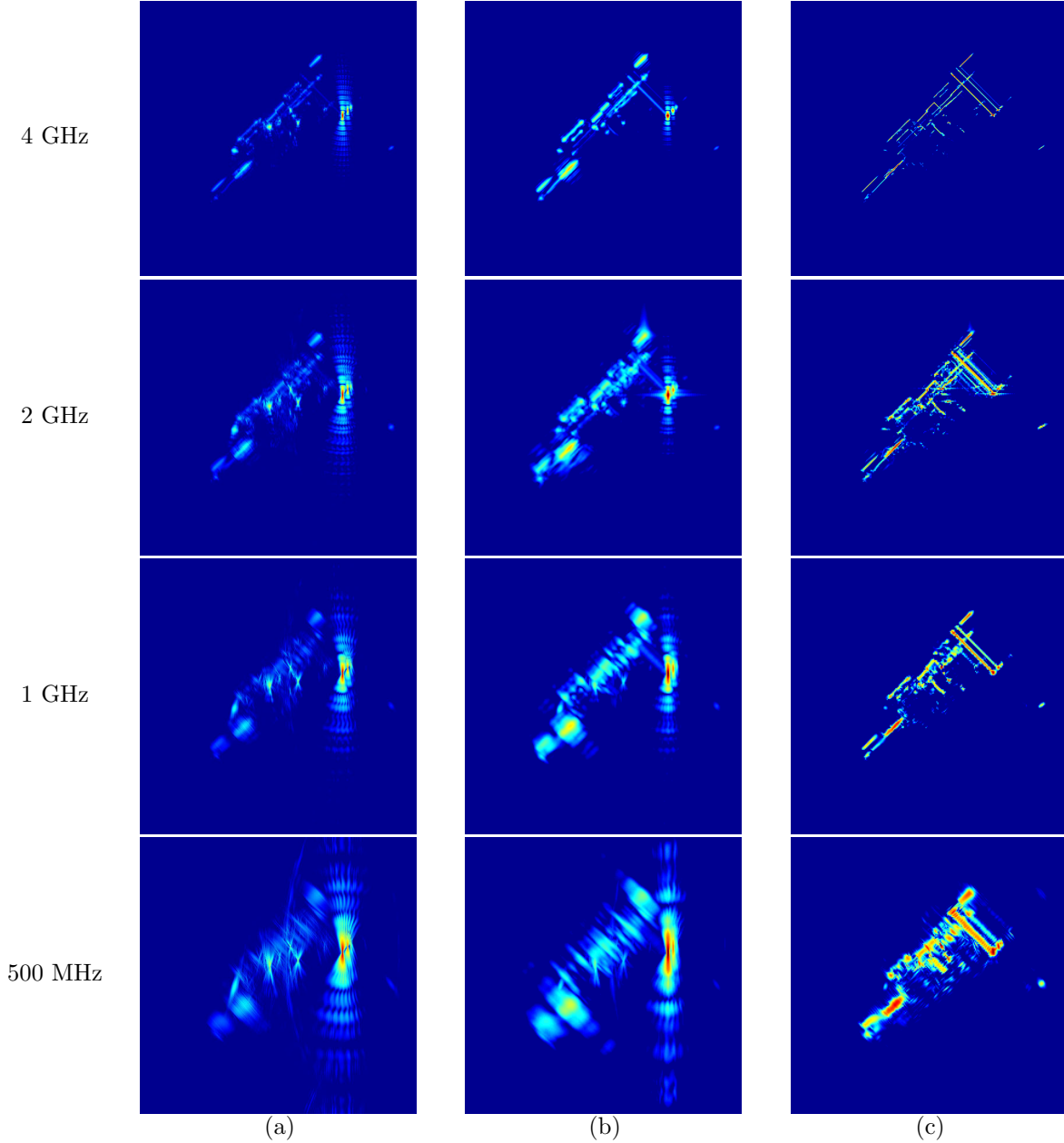**Figure 10.** SAR images of the backhoe with frequency-band omissions (30% of the full-band data available) using a 0°-elevation aperture and bandwidths of 4 GHz, 2 GHz, 1 GHz, and 500 MHz. (a) Conventional imaging. (b) Composite imaging. (c) Composite, point-enhanced imaging.

REFERENCES

1. R. L. Moses, L. Potter, and M. Çetin, "Wide angle SAR imaging," in *Algorithms for Synthetic Aperture Radar Imagery XI*, E. G. Zelnio and F. D. Garber, eds., *Proc. SPIE*, (Orlando, FL, USA), Apr. 2004.
2. M. Çetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Trans. Image Processing* **10**, pp. 623–631, Apr. 2001.
3. "Backhoe Data Sample & Visual-D Challenge Problem available through the Air Force Research Laboratory Sensor Data Management System Web Page: https://www.sdms.afrl.af.mil/main.htm."
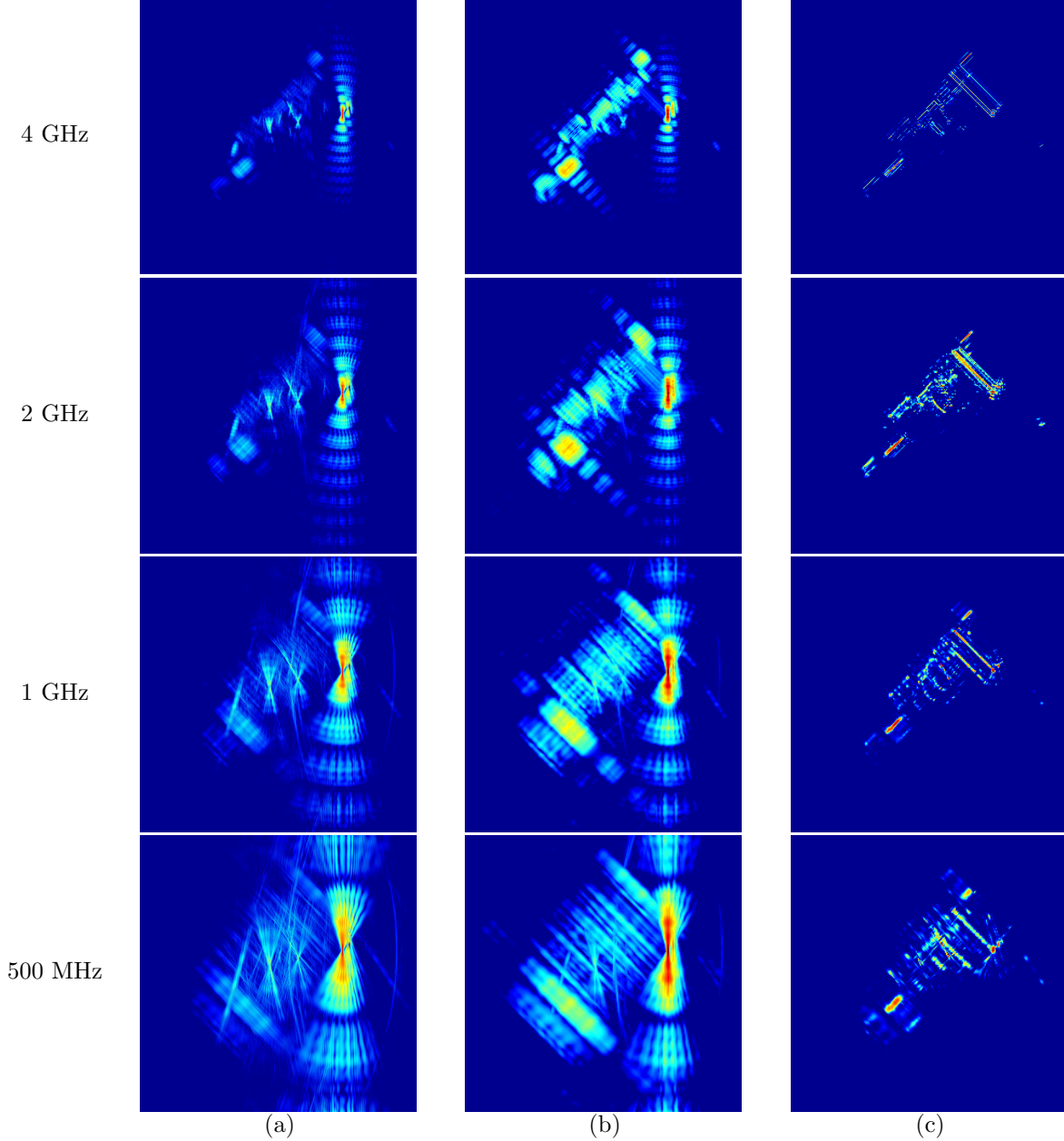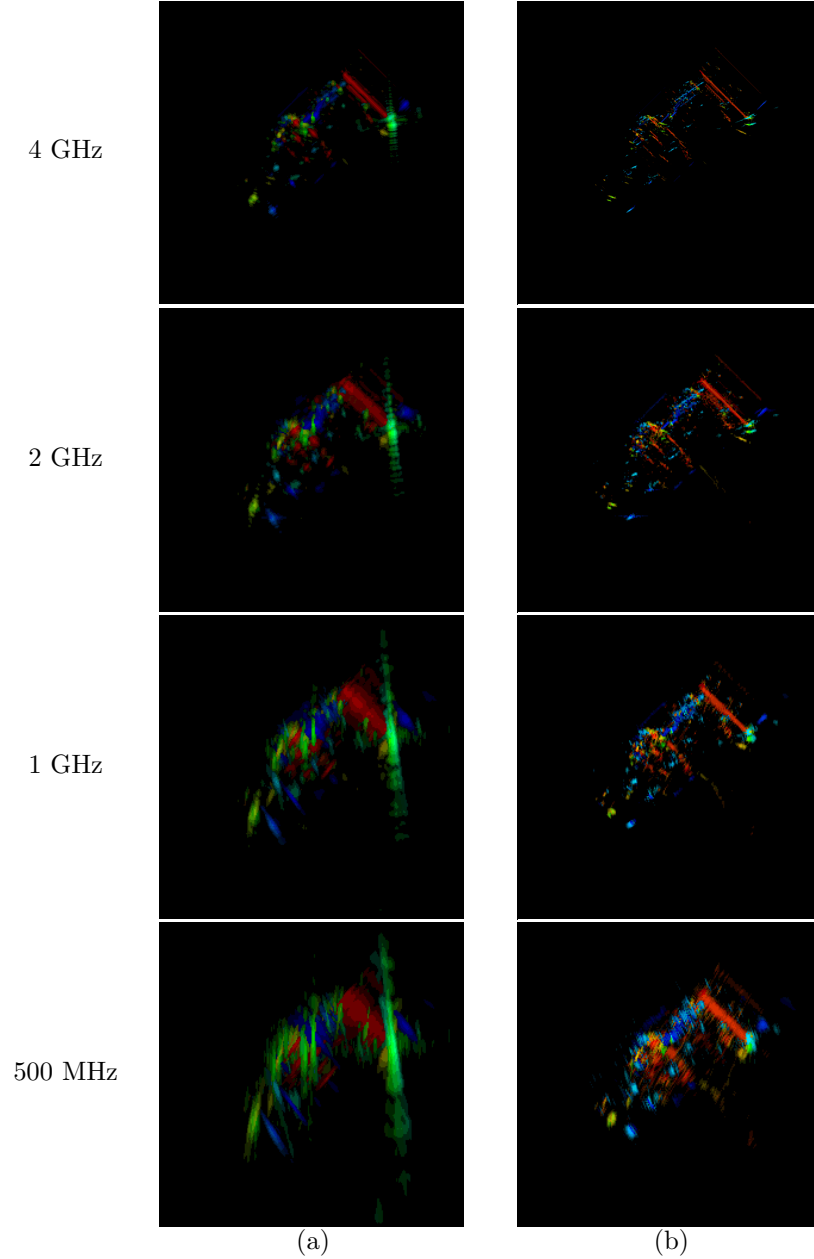
Figure 11. Visualization of aspect dependence. Angle-encoded SAR images of the backhoe with frequency-band omissions (70% of the full-band data available) using a linear aperture and bandwidths of 4 GHz, 2 GHz, 1 GHz, and 500 MHz. (a) Composite imaging. (b) Composite, point-enhanced imaging.

4. M. R. Allen and L. E. Hoff, "Wide-angle wideband SAR matched filter image formation for enhanced detection performance," in *Algorithms for Synthetic Aperture Radar Imagery*, **2230**, pp. 302–314, SPIE, (Orlando, FL, USA), Apr. 1994.

5. R. D. Chaney, A. S. Willsky, and L. M. Novak, "Coherent aspect-dependent SAR image formation," in *Algorithms for Synthetic Aperture Radar Imagery*, **2230**, pp. 256–274, SPIE, (Orlando, FL, USA), Apr. 1994.

Region-enhanced passive radar imaging

M. Çetin and A.D. Lanterman

Abstract: The authors adapt and apply a recently-developed region-enhanced synthetic aperture radar (SAR) image reconstruction technique to the problem of passive radar imaging. One goal in passive radar imaging is to form images of aircraft using signals transmitted by commercial radio and television stations that are reflected from the objects of interest. This involves reconstructing an image from sparse samples of its Fourier transform. Owing to the sparse nature of the aperture, a conventional image formation approach based on direct Fourier transformation results in quite dramatic artefacts in the image, as compared with the case of active SAR imaging. The region-enhanced image formation method considered is based on an explicit mathematical model of the observation process; hence, information about the nature of the aperture is explicitly taken into account in image formation. Furthermore, this framework allows the incorporation of prior information or constraints about the scene being imaged, which makes it possible to compensate for the limitations of the sparse apertures involved in passive radar imaging. As a result, conventional imaging artefacts, such as sidelobes, can be alleviated. Experimental results using data based on electromagnetic simulations demonstrate that this is a promising strategy for passive radar imaging, exhibiting significant suppression of artefacts, preservation of imaged object features, and robustness to measurement noise.

1 Introduction

Traditional synthetic aperture radar (SAR) systems transmit waveforms and deduce information about targets by measuring and analysing the reflected signals. (Ground-based systems looking at airborne targets are generally referred to as inverse SAR (ISAR); for brevity we just use the term SAR.) The active nature of such radars can be problematic in military scenarios since the transmission reveals both the existence and the location of the transmitter. An alternative approach is to exploit 'illuminators of opportunity' such as commercial television and FM radio broadcasts. Such passive approaches offer numerous advantages. The overall system cost may be cheaper, since a transmitter is no longer needed. Commercial transmitters are typically much higher in elevation than the prevailing terrain, yielding coverage of low altitude targets. Most importantly, such a system may remain covert, yielding increased survivability and robustness against deliberate directional interference. Such passive multistatic radar systems, such as Lockheed Martin's Silent Sentry, have been developed to detect and track aircraft. If one could additionally form images from such data, that would be useful in identifying the observed aircraft through image-based target recognition. This provides an alternative to the radar cross-section signature-based automatic target recognition (ATR) method proposed in [1]. Imaging methods are of interest in their own right beyond the ATR application, since a system may encounter targets that are not present in the ATR system's library; in such cases, it would be good to have an image to present to a human analyst. Recently there has been some interest in image reconstruction from passive radar data. In particular, [2] contains a study of the application of well-known deconvolution techniques to passive radar data. The work in [3, 4] proposes the use of time–frequency distributions for passive radar imaging. Finally, [5] contains a derivation of Cramér–Rao bounds for target-shape estimation in passive radar.

Television and FM radio broadcasts operate at wavelengths that are much larger than those typically employed in active radar imaging systems. For instance, an X-band radar might operate at 10 GHz, whereas a passive radar system operates in the VHF and UHF bands (55–885 MHz). From an imaging viewpoint, lower frequencies result in reduced crossrange resolution; hence, to achieve high-resolution images, the target needs to be tracked for some length of time to obtain data over a wide range of angles. Another consequence is that low-frequency images contain extended features, and are not well-modelled by a small number of scattering centres. Furthermore, the signals involved in such broadcasts have much lower bandwidth than the signals used in active radar systems. As a result, given one transmitter–receiver pair, the achievable range resolution is very poor. Hence one needs to make use of multiple transmitters for reasonable coverage in the spatial spectrum.

As a result of these constraints and requirements, forming images of aircraft using passive radar systems involves reconstructing an image from sparse and irregular samples of its Fourier transform [2, 6]. The sampling pattern in a particular data collection scenario depends on the locations of the transmitters and the receiver, as well as the flight path of the object to be imaged; hence it is highly variable. Conventional Fourier transform-based imaging essentially sets the unavailable (due to the sparse aperture) data samples to zeros. This results in various artefacts in

M. Çetin is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

A.D. Lanterman is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Mail Code 0250, Atlanta, GA 30332, USA

E-mail: mcetin@mit.edu

the formed image, the severity of which depends on the specifics of the data collection scenario.

Motivated by the limitations of direct Fourier transform-based imaging in the context of passive radar, an alternative idea of using a deconvolution technique borrowed from radio astronomy (namely the CLEAN algorithm [7, 8]) has been explored in [2]. However, the results of the study in [2], summarised in Section 4.4, suggest that the CLEAN algorithm does not outperform direct Fourier reconstruction for passive radar imaging for the following reasons. The CLEAN algorithm, as well as other deconvolution algorithms based on similar sparse image assumptions, work best on images that are well-modelled as a set of distinct point scatterers. Hence, such algorithms are well-suited to high-frequency imaging of man-made targets, as the current on the scatterer surface tends to collect at particular points. When using low frequencies of interest in passive radar, the images are more spatially distributed. In addition, the complex-valued, and potentially random-phase [9] nature of radar imaging also presents a complication for CLEAN. The complex-valued characteristics of both the underlying image and the observation model produce constructive and destructive interference effects that conspire to obscure true peaks in the underlying reflectance, causing them to be missed by the CLEAN algorithm, and more damagingly create spurious apparent peaks which mislead the algorithm.

To address these challenges we adapt and use a recently-developed, optimisation-based SAR imaging method [10]. This approach uses an explicit model of the particular data collection scenario. This model-based aspect provides significant reduction in the types of artefacts observed in conventional imaging. More importantly, the optimisation framework contains nonquadratic constraints for region-based feature enhancement, which in turn results in accurate reconstruction of spatially extended features. Finally, this approach explicitly deals with the complex-valued and potentially random-phase nature of radar signals. We present experimental results on data obtained through electromagnetic simulations via the Fast Illinois Solver Code (FISC), demonstrating the effectiveness of the proposed approach for passive radar imaging.

2 Data collection in passive radar

In a bistatic radar the transmitter and receiver are at different locations. The angle between the vector from the target to the transmitter and the vector from the target to the receiver, corresponding to the incident and observed directions of the signal, is called the bistatic angle β. For monostatic radar, the bistatic angle is $0°$. Figure 1a illustrates the bistatic radar configuration. The complex-valued data collected at transmitting frequency f is a sample of the Fourier transform of the target reflectivity, and is equivalent to a monostatic measurement taken at the bisecting direction and at a frequency of $f \cos(\beta/2)$ [11, 12]. In a polar co-ordinate system, the bisecting direction gives the azimuthal co-ordinate in Fourier space, and $(4\pi f/c) \cos(\beta/2)$ gives the radial co-ordinate, where c is the speed of light. As the receiver rotates away from the transmitter the bistatic angle β increases and the equivalent frequency $f \cos(\beta/2)$ decreases. When β is $180°$, the measurement is a sample located at the origin in Fourier space. Measurements collected from a receiver that rotates $360°$ around the target lie on a circle in Fourier space, passing through the origin. The diameter of the circle is $4\pi f/c$. Different incident frequencies give data on circles in Fourier space with different diameters, as shown in Fig. 1b. If the transmitter rotates around the target, the circle in Fourier space also rotates by the same amount

and we get more circles of data in Fourier space. Figure 1b illustrates the type of Fourier space coverage obtained through angular and frequency diversity in a bistatic radar.

Unlike the case in active radar systems where one uses high-bandwidth signals, in passive radar based on radio and television signals, one is limited to much lower bandwidths. FM radio has a usable bandwidth of around 45 kHz, and although analogue TV technically has a bandwidth of 6 MHz, little of that is usable for radar purposes. The synchronisation (sync) pulses inherent in the analogue TV signal result in extreme range ambiguities if one attempts traditional matched filtering range compression, as first discovered by Griffiths and Long in the mid-1980s [13]. By the time the signal reaches the receiver, the only significant usable signal is the TV carrier itself, which contains around 50% of the total power in the analogue TV signal (see pp. 20, 21 of [14]). (Having so much power in the carrier may seem wasteful from the standpoint of modern communications, but remember that at the time analogue TV standards were developed the receiver hardware had to be exceedingly simple. Essentially, the transmitter needs to provide its own 'local oscillator' to the receiver.) We can essentially model the usable TV signal as a simple sinusoid. Consequently, at each observation instant, we might think of each transmitter–receiver pair providing essentially 'one point' in the 2-D frequency spectrum. A multistatic system exploiting multiple television and radio stations should be used for obtaining the frequency diversity needed for reasonable quality imaging. The bistatic imaging principle illustrated in Fig. 1 applies to each transmitter/receiver pair in a multistatic system. The aircraft must be tracked and data collected over time to obtain angular diversity, with each transmitter–receiver pair providing data on an arc in 2-D Fourier space. Different transmitters use different frequencies and are at different locations, which leads to multiple arcs of Fourier data, providing further data diversity. In the passive radar scenario explored in this paper, there are multiple transmitters but just one receiver, although the basic idea could easily be expanded to include multiple receivers if appropriate data links are available.

In active synthetic aperture radar, either monostatic or bistatic, one conventional image formation technique is to interpolate the data to a rectangular grid, followed by an inverse Fourier transform. Fourier points outside of the available data support are simply set to zero. In monostatic SAR this is called the polar format algorithm [15–17]. The bistatic version is similar, except the data are placed on the grid with the $\cos(\beta/2)$ warping described above [12, 17]. We can consider a similar approach as the 'conventional' method for imaging in passive radar. In active monostatic radar imaging, the data in the spatial frequency domain usually lie in a regular annular region. The regularity of this region then leads to a sinc-like point spread function when the image is formed using a Fourier transform. On the other hand, in multistatic passive radar, the 'sampling pattern' in the spatial frequency domain is much more irregular for a number of reasons. First, since the transmitted signals are narrowband, each transmitter–receiver pair provides a 'point' rather than a 'slice' of data. Secondly, to obtain reasonable azimuth resolution, data are collected over a wider range of observation angles. Thirdly, the look-angles of different transmitter–receiver pairs lead to coverage in different areas of the spectrum. In a related fashion, where the data lie in the spectrum depends on the flight path of the object being imaged. As a result, when we form images using direct Fourier inversion the imaging artefacts that we encounter are more severe than in the case of active radar
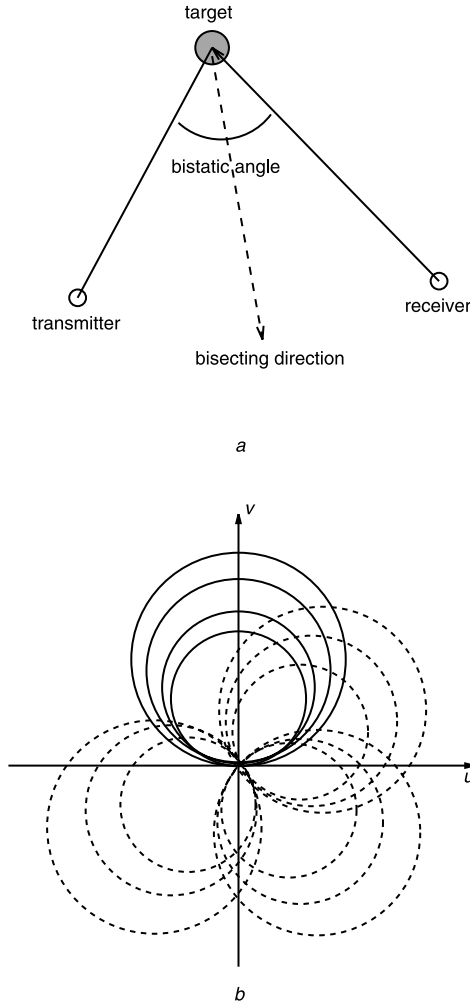
Fig. 1 *Bistatic radar*

a Basic configuration
b Bistatic Fourier space coverage due to angular and frequency diversity
The authors would like to thank Yong Wu, who created these figures for a
DARPA annual report while a student at the University of Illinois

systems. Furthermore, the nature of the artefacts cannot be determined just based on the system design, since the flight path of the aircraft has a role as well.

3 Region-enhanced passive radar imaging

Based on the issues outlined in the previous Section, we propose a different approach for passive radar imaging. Two main ingredients of this approach make it especially suited for passive radar applications. First, it is model-based, meaning that it explicitly uses a mathematical model of the particular observation process. As a result, it has a chance of preventing the types of artefacts that are caused by direct Fourier inversion. Secondly, it facilitates the incorporation of prior information or constraints about the nature of the scenes being imaged. This is important, since passive radar imaging is inherently an ill-posed problem. In particular, we focus on the prior information that at the low frequencies of interest in passive radar, the scenes contain spatially extended structures, corresponding to the actual contours of real aircraft. As a result, we incorporate constraints for preserving and enhancing *region-based* features, such as object contours.

The approach we use for passive radar imaging is based on the feature-enhanced image formation framework of [10], which is built on nonquadratic optimisation. This approach has previously been used in active synthetic aperture

radar imaging. Let us provide a brief overview of feature-enhanced imaging, starting from the following assumed discrete model for the observation process:

$$g = Tf + w \qquad (1)$$

where g denotes the observed passive radar data, f is the unknown sampled reflectivity image, w is additive measurement noise, all column-stacked as vectors, and T is a complex-valued observation matrix. The data can be in the spatial frequency domain, in which case T would be an appropriate Fourier transform-type operator corresponding to the particular sampling pattern determined by the flight path of the target. Alternatively, through a Fourier transform, one can bring the data into the spatial domain, and then use the resulting transformed observations as the input to the algorithm. In this case, T would be the point spread function corresponding to the particular data collection scenario. Our experiments are based on the last-mentioned setup.

The objective of image reconstruction is to obtain an estimate of f based on the data g in (1). Feature-enhanced image reconstruction is achieved by solving an optimisation problem of the following form:

$$\hat{f} = \arg\min_{f}\big\{\|g - Tf\|_2^2 + \lambda_1\|f\|_p^p + \lambda_2\||\nabla|f|\|_p^p\big\} \qquad (2)$$

where $\|\cdot\|_p$ denotes the ℓ_p-norm ($p \le 1$), ∇ is a 2-D derivative operator, $|f|$ denotes the vector of magnitudes of the complex-valued vector f, and λ_1, λ_2 are scalar parameters. The first term in the objective function of (2) is a data fidelity term. The second and third terms incorporate prior information regarding both the behaviour of the field f, and the nature of the features of interest in the resulting reconstructions. The optimisation problem in (2) can be solved by using an efficient iterative algorithm [10], based on half-quadratic regularisation [18]. We describe a basic version of this algorithm in the Appendix.

Each of the last two terms in (2) is aimed at enhancing a particular type of feature that is of importance for radar images. In particular, the term $\|f\|_p^p$ is an energy-type constraint on the solution, and aims to suppress artefacts and increase the resolvability of *point* scatterers. The $\||\nabla|f|\|_p^p$ term, on the other hand, aims to reduce variability in homogeneous *regions*, while preserving and enhancing region boundaries. The relative magnitudes of λ_1 and λ_2 determine the emphasis on such *point-based* against *region-based* features. Therefore this framework lets us reconstruct images with two different flavours: using a relatively large λ_1 yields *point-enhanced* imagery, and using a relatively large λ_2 yields *region-enhanced* imagery. In the context of passive radar imaging, our primary focus is to preserve and enhance the shapes of spatially-distributed objects. Hence we emphasise the use of the region-enhancement terms here.

4 Experiments

4.1 Electromagnetic simulation using FISC

Asymptotic codes such as XPATCH [19] do not work well for aircraft-sized targets at the low frequencies of interest in passive radar systems. Hence, the simulations in the remaining sections invoke the Fast Illinois Solver Code (FISC) [20, 21], which solves Maxwell's equations with the method of moments. FISC is extremely particular about the quality of CAD models it needs. In particular, FISC requires that each edge of each triangular facet exactly match the edge of some other triangular facet. The model must contain no internal or intersecting parts. Unfortunately such models are rare; in particular, readily available models
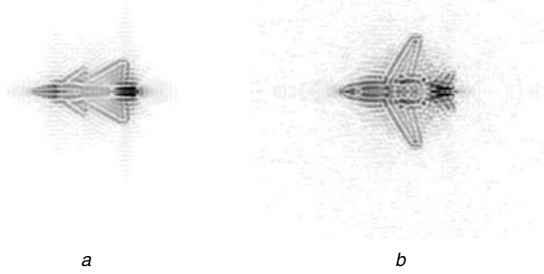
25

Fig. 2 *Reference 256 × 256 passive radar images reconstructed from 'full' datasets using direct Fourier reconstruction*

a VFY-218
b Falcon 20

which are perfectly adequate for XPATCH are often not suitable for FISC.

Each experiment in this paper is conducted on two different targets: a VFY-218, and a Dassault Falcon 20. A FISC compatible model of the VFY-218 comes standard as part of the SAIC Champaign XPATCH/FISC distribution. For the Falcon 20, we started with a Falcon 100 model purchased from Viewpoint Datalabs (now called Digimation), which happened to be FISC compatible. The Falcon 20 is essentially a larger version of the Falcon 100, so we used an approximate Falcon 20 model (as done in [2]) by scaling the Falcon 100 model.

Given such models we construct Fourier datasets through FISC runs. In our experiments we use only the HH-polarisation data. The support of the data in the spatial frequency domain will in general be limited by the observation geometry and system parameters. However, to establish an 'upper bound' on the expected imaging performance, let us first present the images we would obtain if we had a 'full' dataset. To this end, let us use the Fourier data corresponding to 211.25 MHz (NTSC television channel 13) and incident and observed angles over the full 360° viewing circle. Such data would cover a disc in the spatial frequency domain [2]. The magnitudes of the radar images of the two targets, created by inverse Fourier transforming such data, are shown in Fig. 2. Of course, such rich data sets would be unavailable in practice. However, these reconstructions can serve as 'reference scenes' with which to compare the results of our experiments in the following Sections, which are based on realistic data collection scenarios.

4.2 Experimental setup

Figure 3 shows the locations of some high-power VHF television and FM radio stations in the Washington, DC area that are used in our simulations. The centre of the co-ordinate system, where our hypothetical receiver is located, is the Lockheed Martin Mission Systems facility in Gaithersburg, Maryland. Five hypothetical flight paths are shown. The left column of Fig. 4 shows the Fourier 'sampling patterns' resulting from this particular transmitter/receiver geometry for each of the five flight paths. The sampling pattern indicates the support of the observed data in the spatial frequency domain for a particular flight path. Hence, the observed data for each flight path consists of a specific subset of the data used for reconstructing the images of Fig. 2, whose contents are determined by the corresponding sampling pattern. The middle and right columns in Fig. 4 show the magnitude of the corresponding point spread functions (PSFs) given by the inverse Fourier
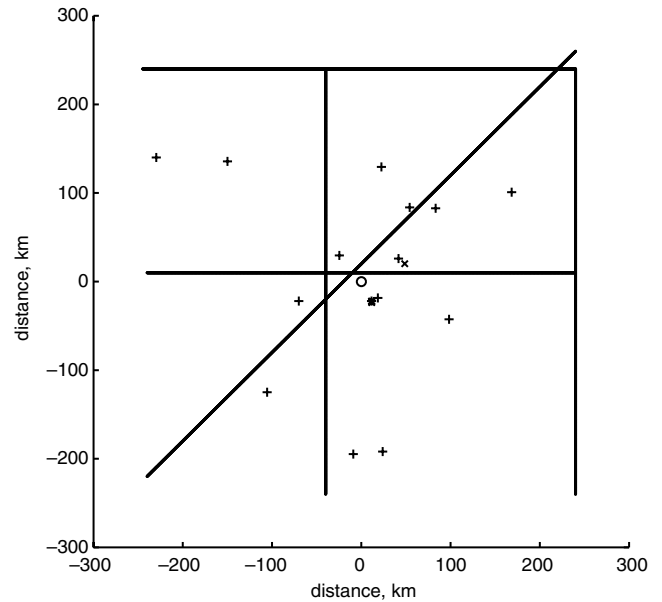

Fig. 3 *Data collection geometry*

VHF TV stations are represented with ×; FM radio stations with +; and receiver with a circle; lines represent five hypothetical flight paths

transform of the sampling patterns. The middle column shows magnitude on a linear scale, while the right column shows magnitude on a logarithmic scale to elucidate low-level detail in the sidelobes. Note that these sampling patterns, or equivalently PSFs, are used in specifying the observation matrix T in (1). The following Section presents results based on data associated with each of these flight paths.

4.3 Region-enhanced imaging results

In all of the experiments presented here, for region-enhanced imaging we use $p = 1$ in (2). For simplicity, we set $\lambda_1 = \lambda_2$ in all examples. This relative parameter choice appears to yield a region-enhanced image, together with suppression of some background artefacts. We choose the absolute values of these parameters based on subjective qualitative assessment of the formed imagery. Automatic selection of these parameters is an open research question. We do not specify the absolute values of λ_1 and λ_2 in the examples we present here, since those numbers are not that meaningful, as they depend on the scaling of the data used.

First consider the flight path corresponding to the sampling pattern in the bottom row in Fig. 4. The corresponding 'conventional' image of the VFY-218, obtained by direct Fourier transformation of the data, is shown in the top row of Fig. 5a. Points in the spatial frequency domain where observations are unavailable are set to zero. This is equivalent to convolving the reference image in Fig. 2a with the PSF in the bottom row of Fig. 4. As compared with the 'reference' image of Fig. 2a, the direct Fourier reconstruction in the top row of Fig. 5a contains severe imaging artefacts, resulting in suppression of some of the characteristic features of the imaged object. In this example we have not added any noise to the measurements. Hence, in the context of the observation model in (1), we do not have any measurement noise. As a result, one can consider applying the pseudoinverse of the observation matrix, namely $T^{\dagger}$, to the data to obtain a reconstruction $\hat{f}_{\mathrm{PINV}} = T^{\dagger}g$. The pseudoinverse reconstruction obtained in this manner is shown in the top row of Fig. 5b. The region-enhanced reconstruction is shown in
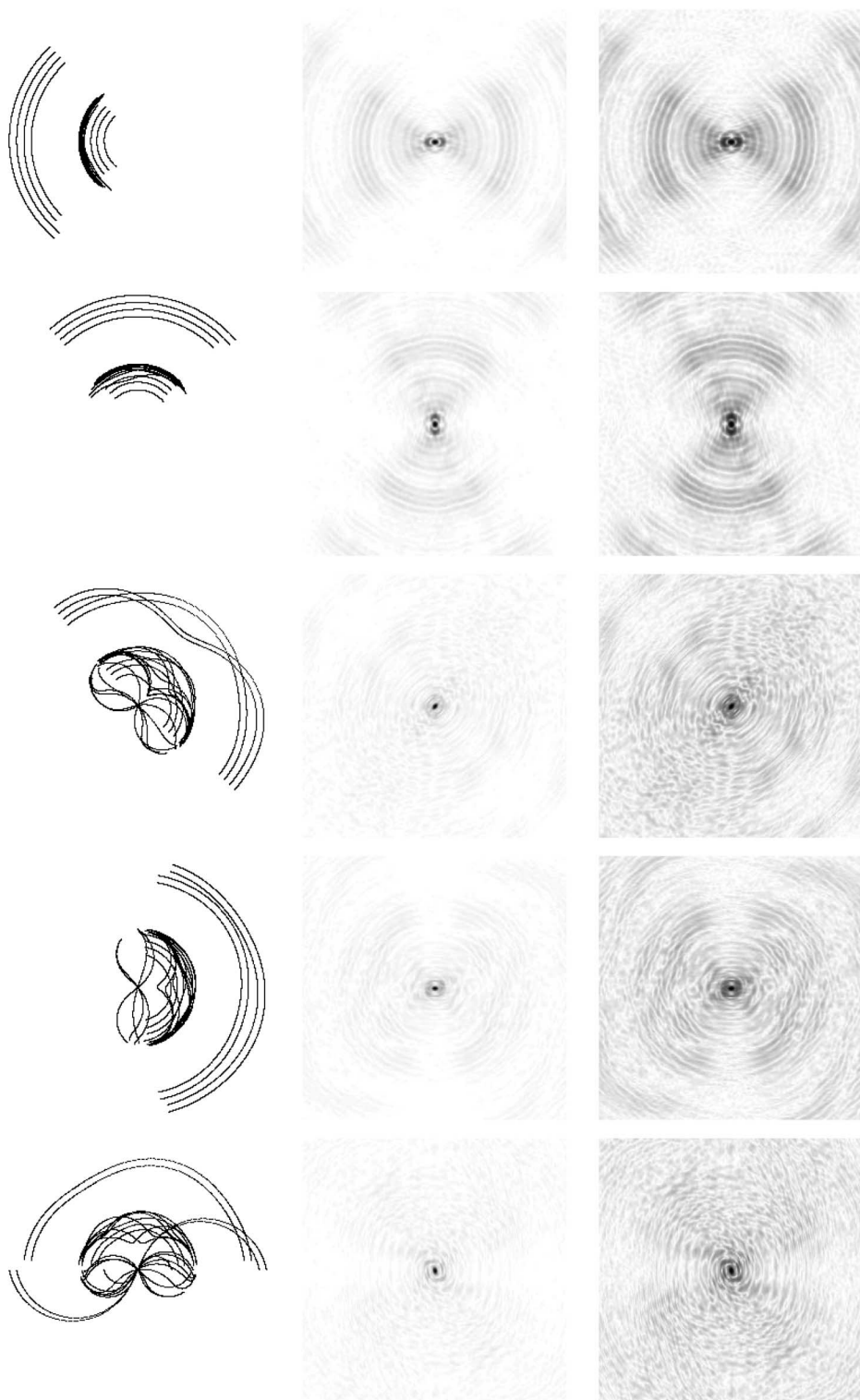
Fig. 4 *Left column shows Fourier sampling patterns associated with five different flight paths; remaining columns show the magnitude of 256 × 256 PSFs associated with sampling patterns; middle column uses linear scale while right column uses logarithmic scale to show fine detail*

the top row of Fig. 5*c*. Both the pseudoinverse and the region-enhanced reconstructions provide reasonable results in this noise-free case, with the region-enhanced reconstruction providing somewhat better suppression of sidelobe artefacts. It is well-known that pseudoinverse solutions are very sensitive to noise, especially when the observation model results in an ill-conditioned matrix. The bottom row of Fig. 5 shows the direct Fourier, the pseudoinverse, and the region-enhanced reconstructions, when we have a small amount of measurement noise. (In these experiments we have added the noise after bringing the data to the spatial domain. Ideally, measurement noise should be added to the phase histories. However, we do not expect that to have any noticeable effect on our results.) The pseudoinverse solution breaks down in this case, and is in general useless in practical scenarios where observation noise is inevitable. The region-enhanced reconstruction exhibits robustness to noise, and preserves the characteristic features and shape of the VFY-218, despite the noisy sparse-aperture observations.
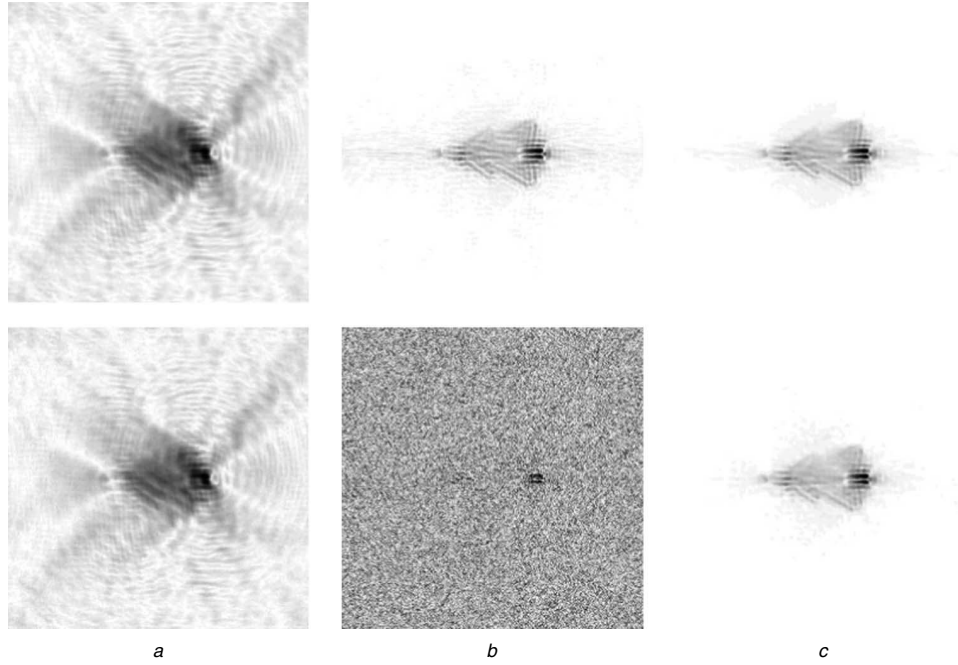
Fig. 5 *Reconstructions of VFY-218 based on data restricted to Fourier sampling pattern shown in bottom row of* Fig. 4

Top row: noiseless data; bottom row: noisy data
a Direct Fourier reconstruction
b Pseudoinverse reconstruction
c Region-enhanced reconstruction

Let us now consider all the flight paths in Fig. 4. In Fig. 6 we show the reconstructions for the VFY-218. In columns (*a*) and (*b*) we have a small amount of measurement noise, resulting in a signal-to-noise ratio (SNR) of 30 dB. (This should be interpreted as an average SNR, since data points may differ in power, yet the measurement noise on each data point has the same variance.) Figures 6*a* and *b* contain the direct Fourier, and the region-enhanced images, respectively. There is a row-to-row correspondence between Figs. 4 and 6, in terms of the flight paths. We observe that region-enhanced imaging produces reconstructions that preserve the features of the reference image of Fig. 2*a* in a much more reliable way than direct Fourier imaging. In columns (*a*) and (*b*) of Fig. 7, we show our results for the Falcon 20, again with data having an SNR of 30 dB, where we can make similar observations to the VFY-218 case. In columns (*c*) and (*d*) of Figs. 6 and 7, we show reconstructions of the VFY-218 and the Falcon 20 respectively, for a noisier scenario where SNR = 10 dB. Region-enhanced imaging appears to produce reasonable results in this case as well.

We also observe that the direct Fourier images in the bottom three rows of Figs. 6 and 7, while blurry, are clearer than the images in the top two rows. Looking at the corresponding sampling patterns in Fig. 4, the primary difference seems to be that the paths corresponding to the top two rows keep the receiver and the transmitters on the same side of the target, yielding a quasi-monostatic (small bistatic angle) geometry, whereas in the bottom three rows, the target flies between the receiver and some of the transmitters, yielding large bistatic angles and wider effective coverage in frequency space. There are two important notes here:

(i) The nature of the artefacts that may be caused by direct Fourier imaging depends on the flight path of the target being imaged, and hence may not be easily predicted prior to data collection. On the other hand, in Figs. 6 and 7 we

observe that region-enhanced images corresponding to different flight paths are much more similar to each other. (ii) The paths where the target crosses between the transmitter and receiver, which give the best performance with conventional direct Fourier reconstruction in our simple simulation as shown in the bottom three rows of Figs. 6 and 7, would be extraordinarily difficult to make work in practice. The direct signal from the transmitter is orders of magnitude larger than the reflected path. Passive radar systems usually alleviate this problem by placing the transmitter in an antenna null (either due to the physical shape of the antenna, or using adaptive nulling techniques in the case of an electronically beamformed array), and maybe also employing some additional RF cancellation techniques. Even with such techniques, the dynamic range requirements are stressing. It would be quite challenging to simultaneously null the direct path signal and receive the reflected signal from an aircraft that is close to the transmitter in angle. For most practical systems, it would be desirable to stick with the quasi-monostatic 'over the shoulder' geometry exemplified by the top two rows of Figs. 4, 6 and 7. Therefore it is important to have a technique like region-enhanced imaging which can generate reasonable images in such quasi-monostatic scenarios.

On a laptop PC with a 1.80 GHz Intel Pentium-4 processor, the average computation time for the region-enhanced images presented (each composed of 256×256 pixels) was around 100 seconds, using non-optimised MATLAB code.

Finally, let us test the robustness of this image formation technique to an extreme amount of measurement noise. In Fig. 8, we consider a scenario where SNR $= -10$ dB, and for the sake of space, we consider only one of the objects, namely the VFY-218, and only one of the flight paths, namely the one in the bottom row of Fig. 4. The conventional image in Fig. 8*a* is dominated by noise artefacts. On the other hand, the region-enhanced image in

Fig. 6 *Reconstructions of VFY-218 based on data restricted to Fourier sampling patterns shown in Fig. 4*

a Direct Fourier reconstructions, SNR = 30 dB
b Region-enhanced reconstructions, SNR = 30 dB
c Direct Fourier reconstructions, SNR = 10 dB
d Region-enhanced reconstructions, SNR = 10 dB

Fig. 8*b* preserves the basic shape of the aircraft, despite some degradation in the image due to noise.

4.4 Experiments with CLEAN

To illustrate the need for a sophisticated technique like the region-enhanced approach used in the previous Section we conclude our experiments with some results using a simple CLEAN algorithm [7]. In the CLEAN algorithm, one finds the point with the largest magnitude in the 'dirty map' (i.e. the conventional direct Fourier transform reconstruction) to be CLEANed, shifts the PSF of the system to that point, and normalises the PSF so that its origin equals the value of the image at the found peak multiplied by a parameter called the 'loop gain'. This shifted and normalised PSF is subtracted from the dirty map. A single point, corresponding to where the peak was in the dirty map, is added to a 'clean map' which is built up as the algorithm
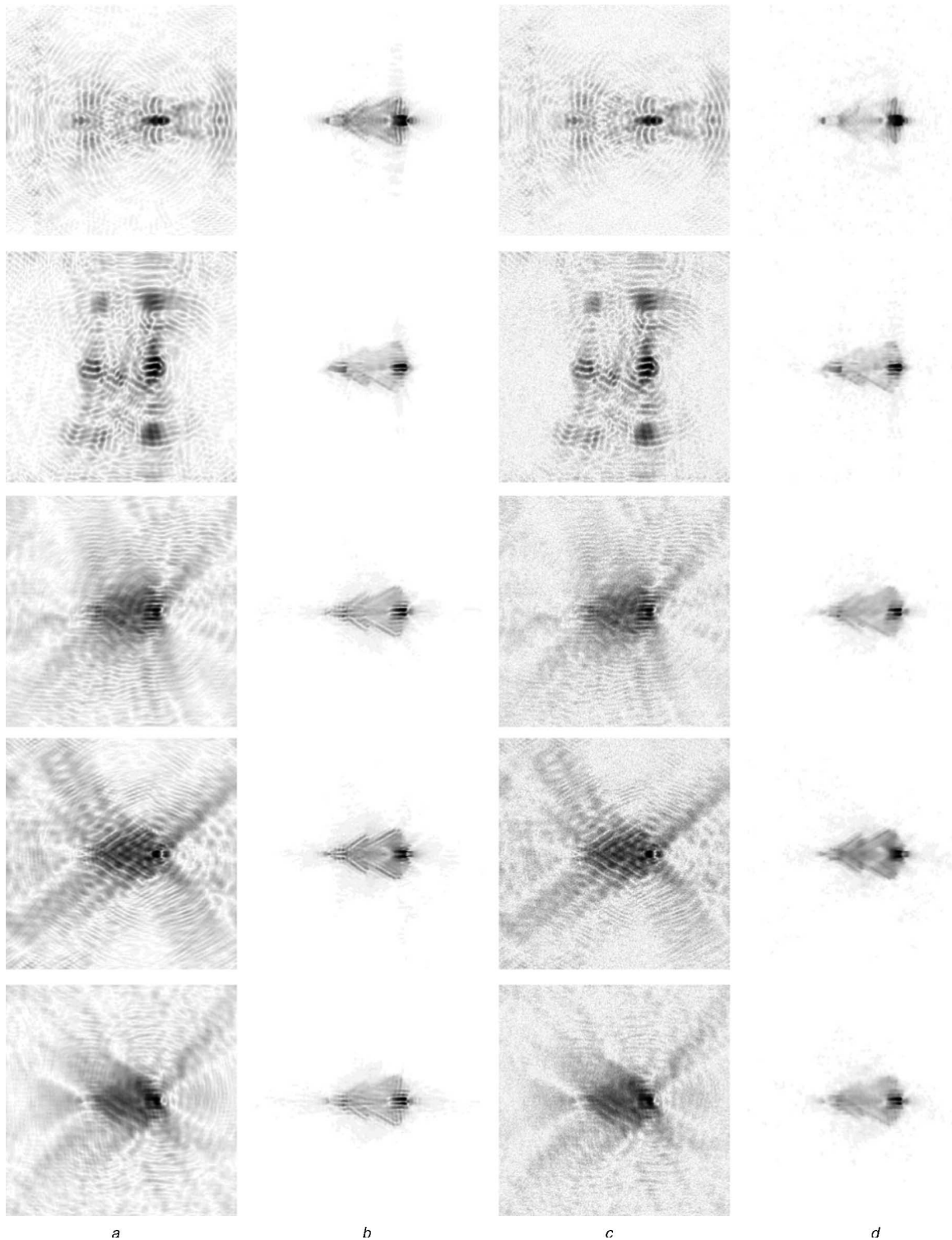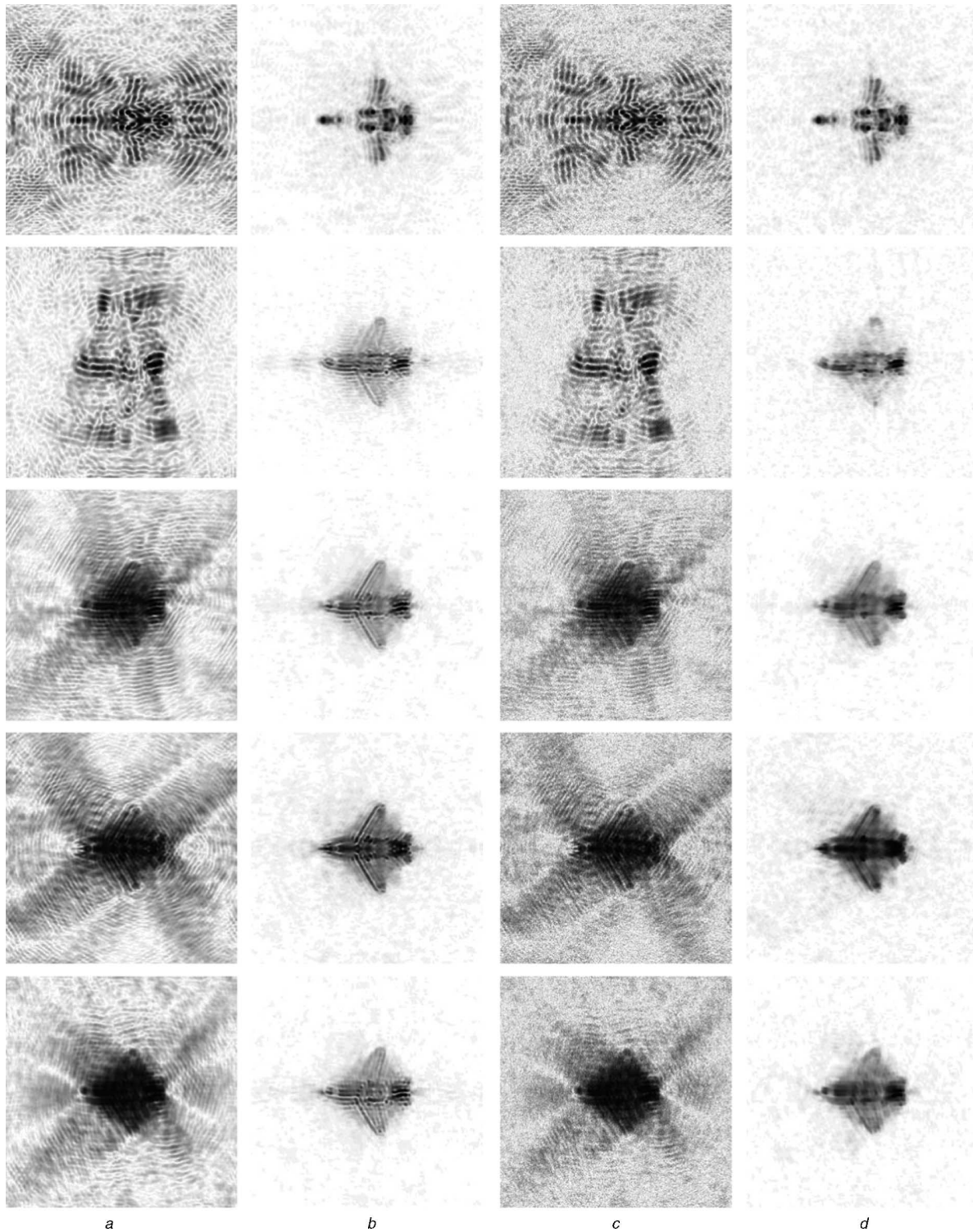
Fig. 7 *Reconstructions of Falcon 20 based on data restricted to Fourier sampling patterns shown in Fig. 4*

a Direct Fourier reconstructions, SNR = 30 dB
b Region-enhanced reconstructions, SNR = 30 dB
c Direct Fourier reconstructions, SNR = 10 dB
d Region-enhanced reconstructions, SNR = 10 dB

proceeds. The procedure is iterated until some stopping criterion is met.

Figure 9 shows the results of 400 iterations of the CLEAN algorithm on the VFY-218 and the Falcon 20, based on noiseless data. (The raw CLEAN images are sparse and may be difficult to reproduce in print in their original state. Hence, the magnitudes of the radar images have been blurred by a Gaussian kernel, and the images are displayed on a square-root scale to make sure that

faint features appear after copying.) We use a loop gain of 0.15, which has been a typical choice in radio astronomy applications of CLEAN. Again, there is a row-to-row correspondence between Figs. 4 and 9 in terms of the flight paths. These results should be compared with those of direct Fourier reconstruction and region-enhanced imaging in Figs. 6 and 7. Although CLEAN has excelled in a number of high-resolution imaging scenarios, it does not seem to outperform standard direct Fourier
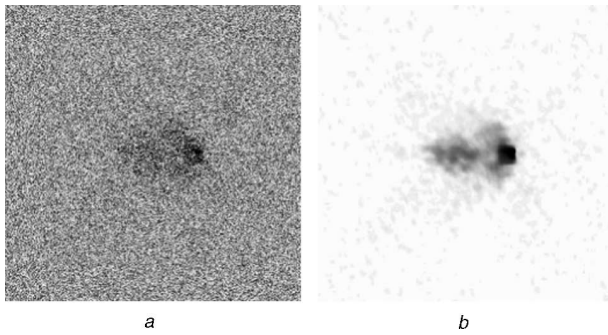
Fig. 8 *Reconstructions of VFY-218 based on data (with SNR = −10 dB) restricted to Fourier sampling pattern shown in bottom row of Fig. 4*

a Direct Fourier reconstruction
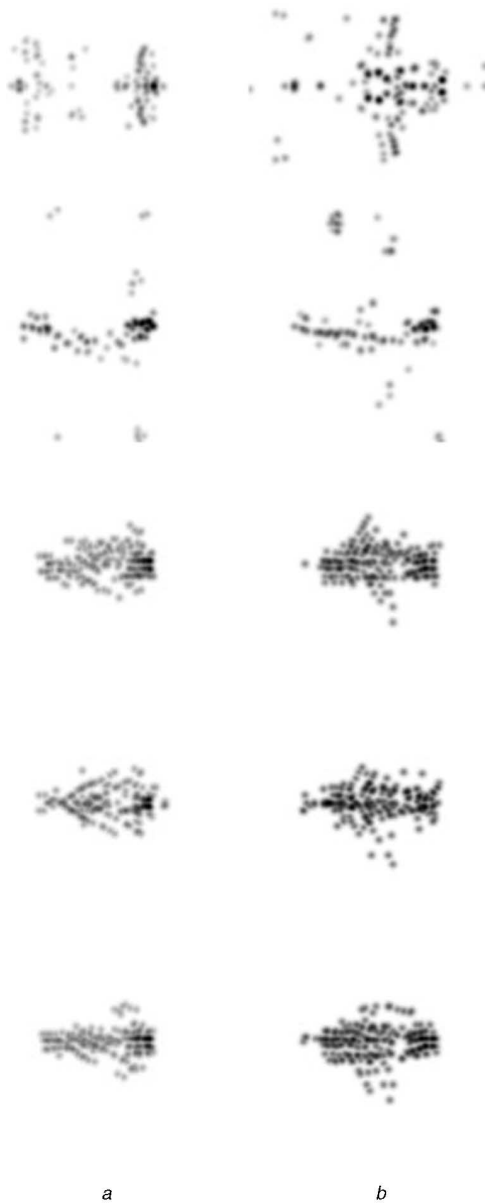*b* Region-enhanced reconstruction

Fig. 9 *Results of 400 iterations of CLEAN algorithm on noiseless data with loop gain of 0.15*

a VFY-218
b Falcon 20

reconstruction in the context of passive radar imaging. On the other hand, region-enhanced imaging appears to provide significantly improved imagery as compared to both Fourier reconstruction and CLEAN.

5 Limitations and possible extensions

In this paper we have assumed that the direct signal from the transmitter is available to provide a phase reference for the reflected signal from the target. More problematically, we have assumed that we know the passive radar observation model exactly, which involves knowledge about not only the transmitters and the receiver, but also about the flight path of the target being imaged. In practice, information about the target flight path is obtained from a tracking system, and will contain uncertainties. The uncertainties in the estimated path will be manifest as phase errors in the data. Considering that the phase of the Fourier transform of an image contains significant information, it is important to develop image formation techniques that can deal with such uncertainties in the observation model. The SAR community refers to such techniques as autofocus algorithms [17, 22]. Such an extension of the image formation technique we presented constitutes a challenging direction for future work. Maneuvering targets that may be rolling, pitching, and yawing in complex ways would present further challenges, even if the target positions over time were exactly known.

Our imaging model assumes isotropic point scattering. However, when the imaged object is observed over a wide range of angles, the aspect-dependent amplitude of scattering returns can become significant. Performing region-enhanced passive radar imaging under aspect and/or frequency-dependent anisotropic scattering would be an interesting extension of our work. Along these lines, the use of time–frequency transforms for wide-angle imaging, motivated by the passive radar application, is discussed in [3] although its authors do not explicitly discuss how to address sparse apertures.

Our final remark is on frequency-dependent scattering. The tomographic radar model [12, 16] suggests that bistatic data at one frequency can be used to synthesise data at multiple lower frequencies. This assumption of frequency-independent scattering was employed in two places in our paper. It was used both in the construction of the observation model, and also in the creation of the simulated data. Since FISC runs are computationally expensive, we took advantage of this assumption and conducted a single run at 211.25 MHz. The fidelity of our simulations could be improved by conducting appropriate separate FISC runs for all the transmitters employed, even if no changes are made to the model used to form images from the data. A good avenue for future work would be to find out how far one could push the underlying bistatic equivalence theorems [23–25] in simulating data, before the disadvantage of lost accuracy due to frequency-dependent scattering exceeds the advantage of shorter computation times.

6 Conclusions

We have explored the use of an optimisation-based, region-enhanced image formation technique for the sparse-aperture passive radar imaging problem. Due to the sparse and irregular pattern of the observations in the spatial frequency domain, conventional direct Fourier transform-based imaging from passive radar data leads to unsatisfactory results, where artefacts are produced and characteristic features of the imaged objects are suppressed. The region-enhanced imaging approach we use appears to be suited to the passive radar imaging problem for a number of reasons. First, due to its model-based nature, the types of artefacts caused by conventional imaging are avoided. Secondly, it leads to the preservation and enhancement of spatially extended object

features. Thirdly, unlike a number of deconvolution techniques, it can deal with the complex-valued nature of the signals involved. Our experimental results based on data obtained through electromagnetic simulations demonstrate the effectiveness and promise of this approach for passive radar imaging.

7 Acknowledgment

8 References

1 Ehrman, L.M., and Lanterman, A.D.: 'Automatic target recognition via passive radar, using precomputed radar cross-sections and a co-ordinated flight model', *IEEE Trans. Aerosp. Electron. Syst.*, submitted in 2003
2 Lanterman, A.D., and Munson, Jr., D.C.: 'Deconvolution techniques for passive radar imaging', in Zelnio, E.G. (Ed.): 'Algorithms for Synthetic Aperture Radar Imagery IX' Orlando, FL, USA, Apr. 2002, (*Proc. SPIE*, **4727**), pp. 166–177
3 Lanterman, A.D., Munson, Jr., D.C., and Wu, Y.: 'Wide-angle radar imaging using time-frequency distributions', *IEE Proc., Radar Sonar Navig.*, 2003, **150**, (4), pp. 203–211
4 Wu, Y., and Munson, Jr., D.C.: 'Multistatic passive radar imaging using the smoothed pseudo Wigner-Ville distribution'. Proc. IEEE Int. Conf. on Image Processing, Oct. 2001, Vol. 3, pp. 604–607
5 Ye, J.C., Bresler, Y., and Moulin, P.: 'Cramér–Rao bounds for 2-D target shape estimation in nonlinear inverse scattering problems with application to passive radar', *IEEE Trans. Antennas Propag.*, 2001, **49**, (5), pp. 771–783
6 Wu, Y., and Munson, Jr., D.C.: 'Multistatic synthetic aperture imaging of aircraft using reflected television signals', in Zelnio, E.G. (Ed.): 'Algorithms for Synthetic Aperture Radar Imagery VIII', Orlando, FL, USA, Apr. 2001, (*Proc. SPIE*, **4382**)
7 Högbom, J.: 'Aperture synthesis with a non-regular distribution of interferometer baselines', *Astron. Astrophy. Suppl. Ser.*, 1974, **15**, pp. 417–426
8 Schwarz, U.J.: 'Mathematical–statistical description of the iterative beam removing technique (method CLEAN)', *Astron. Astrophys.*, 1978, **65**, pp. 345–356
9 Munson, Jr., D.C., and Sanz, J.L.C.: 'Image reconstruction from frequency-offset Fourier data', *Proc. IEEE*, 1984, **72**, (6), pp. 661–669
10 Çetin, M., and Karl, W.C.: 'Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization', *IEEE Trans. Image Process.*, Apr. 2001, **10**, (4), pp. 623–631
11 Mensa, D., and Heidbreder, G.: 'Bistatic synthetic aperture radar imaging of rotating objects', *IEEE Trans. Aerosp. Electron. Syst.*, 1992, **18**, pp. 423–431
12 Arıkan, O., and Munson, Jr., D.C.: 'A tomographic formulation of bistatic synthetic aperture radar'. Proc. of Conf. on Advances in Communications and Control Systems ComCon 88, Oct. 1988
13 Griffiths, H.D., and Long, N.R.W.: 'Television-based bistatic radar', *IEE Proc. F, Commun. Radar Signal Process.*, 1986, **133**, (7), pp. 649–657
14 Howland, P.E.: 'Television based bistatic radar' University of Birmingham, UK, 1997
15 Walker, J.: 'Range-Doppler imaging of rotating objects', *IEEE Trans. Aerosp. Electron. Syst.*, 1980, **AES-16**, (1), pp. 23–52
16 Munson, Jr., D.C., O'Brien, J.D., and Jenkins, W.K.: 'A tomographic formulation of spotlight-mode synthetic aperture radar', *Proc. IEEE*, 1983, **71**, pp. 917–925
17 Jakowatz, Jr., C.V., Wahl, D.E., Eichel, P.H., Ghiglia, D.C., and Thompson, P.A.: 'Spotlight-mode synthetic aperture radar: a signal processing approach' (Kluwer Academic, Norwell, MA, 1996)
18 Geman, D., and Reynolds, G.: 'Constrained restoration and the recovery of discontinuities', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1992, **14**, (3), pp. 367–383
19 Hazlett, M., Andersh, D.J., Lee, S.W., Ling, H., and Yu, C.L.: 'XPATCH: a high-frequency electromagnetic scattering prediction code using shooting and bouncing rays', in Watkins, W.R., and Clement, D. (Eds.): 'Targets and Backgrounds: Characterization and Representation' Orlando, FL, USA, Apr. 1995 (*Proc. SPIE*, **2469**), pp. 266–275
20 Song, J.M., and Chew, W.C.: 'The Fast Illinois Solver Code: requirements and scaling properties', *IEEE Comput. Sci. Eng.*, 1998, **5**, (3), pp. 19–23
21 Song, J.M., Lu, C.C., Chew, W.C., and Lee, S.W.: 'Fast Illinois solver code (FISC)', *IEEE Antennas Propag. Mag.*, 1998, **40**, (3), pp. 27–34
22 Morrison, R.L.: 'Entropy-based autofocus for synthetic aperture radar'. MS thesis, University of Illinois at Urbana-Champaign 2002
23 Glaser, J.I.: 'Bistatic RCS of complex objects near forward scatter', *IEEE Trans. Aerosp. Electron. Syst.*, 1985, **21**, (1), pp. 70–78
24 Glaser, J.I.: 'Some results in the bistatic radar cross-section (RCS) of complex objects', *Proc. IEEE*, 1989, **77**, (5), pp. 639–648
25 Eigel, R.L., Collins, P.J., Terzouli, A.J., Nesti, G., and Fortuny, J.: 'Bistatic scattering characterization of complex objects', *IEEE Trans. Geosci. Remote Sens.*, 2000, **38**, (5), pp. 2078–2092

9 Appendix: Numerical algorithm for region-enhanced imaging

To find a local minimum of the optimisation problem in (2), we use a basic version of the numerical algorithm proposed in [10]. This algorithm is based on ideas from half-quadratic regularisation [18], and can be shown to yield a quasi-Newton scheme with a special Hessian approximation. The algorithm is convergent in terms of the cost functional. In this Section, we only present the most basic form of this algorithm. Our goal here is only to provide a recipe for implementation, rather than a discussion of the properties of this numerical scheme.

To avoid problems due to nondifferentiability of the ℓ_p-norm around the origin when $p \leq 1$, we use the following smooth approximation to the ℓ_p-norm in (2):

$$\|z\|_p^p \approx \sum_{i=1}^{K} (|(z)_i|^2 + \epsilon)^{p/2} \quad (3)$$

where $\epsilon \geq 0$ is a small constant, K is the length of the complex vector z, and $(z)_i$ denotes its ith element. For numerical purposes, we thus solve the following slightly modified optimisation problem:

$$\hat{f} = \arg \min_f \left\{ \|g - Tf\|_2^2 + \lambda_1 \sum_{i=1}^{N} (|(f)_i|^2 + \epsilon)^{p/2} \right.$$

$$\left. + \lambda_2 \sum_{i=1}^{M} (|(\nabla|f|)_i|^2 + \epsilon)^{p/2} \right\} \quad (4)$$

Note that we recover the original problem in (2) as $\epsilon \to 0$. The stationary points of the cost functional in (4) satisfy

$$H(f)f = T^H g \quad (5)$$

where

$$H(f) \triangleq T^H T + \lambda_1 \Lambda_1(f) + \lambda_2 \Phi^H(f) \nabla^T \Lambda_2(f) \nabla \Phi(f) \quad (6)$$

$$\Lambda_1(f) \triangleq \text{diag} \left\{ \frac{p/2}{(|(f)_i|^2 + \epsilon)^{1-p/2}} \right\}$$

$$\Lambda_2(f) \triangleq \text{diag} \left\{ \frac{p/2}{(|(\nabla|f|)_i|^2 + \epsilon)^{1-p/2}} \right\}$$

$$\Phi(f) \triangleq \text{diag} \{ \exp(-j\phi[(f)_i]) \}$$

Here $\phi[(f)_i]$ denotes the phase of the complex number $(f)_i$, $(\cdot)^H$ denotes the Hermitian of a matrix, and $\text{diag}\{\cdot\}$ is a diagonal matrix whose ith diagonal element is given by the expression inside the brackets. Based on this observation, the most basic form of the numerical algorithm we use is as follows:

$$H(\hat{f}^{(n)}) \hat{f}^{(n+1)} = T^H g \quad (7)$$

where n denotes the iteration number. We run the iteration in (7) until $\|\hat{f}^{(n+1)} - \hat{f}^{(n)}\|_2^2 / \|\hat{f}^{(n)}\|_2^2 < \delta$, where $\delta > 0$ is a small constant.

Joint Image Formation and Anisotropy Characterization in Wide-Angle SAR

Kush R. Varshney[a], Müjdat Çetin[a,b], John W. Fisher III[c] and Alan S. Willsky[a,c]

[a]Laboratory for Information and Decision Systems,
Massachusetts Institute of Technology, Cambridge, MA 02139, USA;

[b]Faculty of Engineering and Natural Sciences,
Sabancı University, Orhanlı, Tuzla 34956 İstanbul, Turkey;

[c]Computer Science and Artificial Intelligence Laboratory,
Massachusetts Institute of Technology, Cambridge, MA 02139, USA

ABSTRACT

We consider the problem of jointly forming images and characterizing anisotropy from wide-angle synthetic aperture radar (SAR) measurements. Conventional SAR image formation techniques assume isotropic scattering, which is not valid with wide-angle apertures. We present a method based on a sparse representation of aspect-dependent scattering with an overcomplete basis composed of basis vectors with varying levels of angular persistence. Solved as an inverse problem, the result is a complex-valued, aspect-dependent response for each spatial location in a scene. Our non-parametric approach does not suffer from reduced cross-range resolution inherent in subaperture methods and considers all point scatterers in a scene jointly. The choice of the overcomplete basis set incorporates prior knowledge of aspect-dependent scattering, but the method is flexible enough to admit solutions that may not match a family of parametric functions. We enforce sparsity through regularization based on the ℓ_k-norm, $k < 1$. This formulation leads to an optimization problem that is solved through a robust quasi-Newton method. We also develop a graph-structured interpretation of the overcomplete basis leading towards approximate algorithms using guided depth-first search with appropriate stopping conditions and search heuristics. We present experimental results on synthetic scenes and the backhoe public release dataset.

Keywords: synthetic aperture radar, wide-angle imaging, anisotropy, sparse signal representation, image formation, inverse problems

1. INTRODUCTION

Wide-angle synthetic aperture radar (SAR) imaging has come to the fore recently due to advances in navigation and avionics technologies that permit the synthesis of very long apertures. In principle, wide-angle measurements allow for the formation of images finely resolved in the cross-range direction. However, conventional image formation techniques are not adequate for dealing with data collected over wide-angle apertures for a number of reasons. One issue, and the focus of this paper, that arises with wide-angle apertures is that dependence of scattering behavior on aspect angle, termed *anisotropy*, becomes prominent because objects are viewed from different sides rather than from nearly the same point of view. This is in opposition to narrow-angle imaging, where it is a fairly reasonable assumption that scattering amplitude is constant over the aperture. In conventional image formation techniques, the failure to model angle dependence results in an averaging over that variable, leading to inaccurate scattering estimates. In addition, the anisotropy level of scatterers is not characterized. Yet, anisotropy characterization may be used as a feature for automatic target recognition and for improved image formation.

The problem of detecting, estimating, and modeling aspect-dependent scattering behavior has received attention lately. Anisotropy characterization methods may be broadly categorized into those that operate in the phase history domain, employing parameterizations for angle-dependent scattering, and those that operate in

---

the image domain. The general parametric approach is to posit a parametric model for angle dependent scattering, often motivated by electromagnetic theory, and estimate the model parameters, leading to a well-defined estimation problem.[1–4] Image-domain methods use a multiaperture approach for characterizing anisotropy.[5–11] Subaperture images are formed, either conventionally or using an enhanced image formation technique, from segments of the measurements divided in aspect angle. The sequence of subaperture images then gives an indication of the persistence of the scatterers in the scene. It should be noted that the subaperture images have poorer cross-range resolution than an image formed from the full aperture would. Also, subapertures are of fixed angular extent; consequently, any subaperture analysis is limited in its ability to characterize anisotropy persistence. Parametric methods and image-domain methods have been shown to work well in different situations. Notably, the parametric models incorporate much prior information about expected scattering behavior. Also, the estimated parameters have physical significance, e.g. a parameter corresponding to the physical length of the scattering mechanism. The image-domain methods are robust, easy to reason about conceptually, and can be applied to already formed images.

The parametric model formulation of the anisotropy characterization problem is of course predicated on the correct modeling of natural phenomena. However, parametric models often do not hold in wide-angle imaging scenarios.[10] Within the image-domain methods, a subaperture pyramid framework with overlapping subapertures of various angular extents moves towards allowing a continuum of aspect angle extents, but is still limited to full-, half-, quarter-, ⋯⋯-, apertures.[12] Also, in most techniques, the characterization of anisotropy in different spatial locations (different pixels) is done independently.

In this paper, we consider an inverse problem formulation utilizing an overcomplete basis and sparsifying regularization for joint image formation and anisotropy characterization in wide-angle SAR. Sparsifying regularization has been applied to inverse problems including acoustic source localization[13] and isotropic SAR imaging,[14] but has not previously been applied to the SAR anisotropy characterization problem. While still taking advantage of prior information, this method is flexible enough to admit solutions that are not from a prespecified parametric family. It jointly treats spatial locations and suffers no reduction in cross-range resolution. We also develop a graph-structured interpretation of our overcomplete basis leading towards novel approximate algorithms to solve the inverse problem. These algorithms, having reduced memory requirements, may well find application in a wide variety of sparse signal representation settings beyond the specific problem of anisotropy in SAR.

Sec. 2 describes our framework for bringing the SAR image formation and anisotropy characterization application together with the inverse problem–sparsity mathematical formalism. Specifically, an overcomplete expansion of the point-scattering observation model is proposed, along with a discussion on the choice of vectors for the expansion. In Sec. 3, we build upon the framework of the previous section, describing methods of solving the inverse problem while imposing sparsity. A quasi-Newton method[14] and greedy graph-structured algorithms are applied to the problem. In Sec. 4, examples with synthetic data, with a scene composed of realistic canonical scatterers, and a scene containing the backhoe loader of the Backhoe Data Dome[15] are given. We provide some discussion of the results in the concluding section.

2. OVERCOMPLETE BASIS FORMULATION FOR ANISOTROPY CHARACTERIZATION AND IMAGE FORMATION

In this section, we describe a formulation of the anisotropy characterization problem which differs from the subaperture and parametric formulations mentioned in Sec. 1. The problem is approached by constructing an overcomplete basis and appropriately using the phase history measurements. An overcomplete basis, also known as an overcomplete dictionary, is more than a basis, i.e. a collection containing more vectors than necessary to span the space and hence, a linearly dependent set. The idea is to expand the aspect-dependent scattering function $s(\theta)$ at each spatial location as a superposition of basis vectors and then determine coefficients for those vectors. The first part of the section leaves the overcomplete basis fully general; the section concludes with a consideration of specific basis choices.

2.1. Anisotropy Characterization Inverse Problem

In two-dimensional imaging, the goal is to determine the complex-valued scattering function of a ground patch $s(x, y)$, where x and y are coordinates with origin at the center of that ground patch in the range and cross-range

directions, respectively. However, due to anisotropy, scattering depends on aspect angle θ, with the scattering function taking the form $s(x, y, \theta)$. In our work, we aim to jointly characterize anisotropy and form images by determining this function $s(x, y, \theta)$. The starting point for our overcomplete expansion is the phase history observation model for point scattering centers with anisotropy, given below.

$$r(f, \theta) = \sum_{p=1}^{P} s(x_p, y_p, \theta) \exp \left\{ -j \frac{4\pi f}{c} (x_p \cos \theta + y_p \sin \theta) \right\},$$

where c is the speed of propagation and f is the frequency of the radar measurements. For a single spatial location p, we expand the aspect-dependent scattering as follows:

$$s(x_p, y_p, \theta) = \sum_{m=1}^{M} a_{p,m} b_m(\theta),$$

yielding the following overall $M \cdot P$ vector basis expansion:

$$r(f, \theta) = \sum_{p=1}^{P} \sum_{m=1}^{M} a_{p,m} b_m(\theta) \exp \left\{ -j \frac{4\pi f}{c} (x_p \cos \theta + y_p \sin \theta) \right\}.$$

Isotropic scattering is a special case of the above expression with $M = 1$ and $b_1(\theta)$ constant.

Assuming that the phase history measurements are at K discrete frequencies and N discrete aspect angles, let us define length N vectors $\mathbf{r}_k = r(f_k, \theta)$, $\mathbf{b}_m = b_m(\theta)$, and $\boldsymbol{\varepsilon}_{k,p} = \exp \left\{ -j \frac{4\pi f_k}{c} (x_p \cos \theta + y_p \sin \theta) \right\}$. Then, taking $\boldsymbol{\phi}_{k,p,m} = \mathbf{b}_m \boldsymbol{\varepsilon}_{k,p}$, the basis expansion may be simply expressed as:

$$\mathbf{r}_k = \sum_{p=1}^{P} \sum_{m=1}^{M} a_{p,m} \boldsymbol{\phi}_{k,p,m}, \; k = 1, \dots, K. \tag{1}$$

The inverse problem is to determine the $M \cdot P$ complex-valued coefficients $a_{p,m}$ that satisfy or approximately satisfy the linear equations (1). We choose the number of basis vectors M such that $M > N$, making the basis overcomplete.

Now, let us move to matrix-vector equations to simplify the discussion. The collection of all phase history measurements can be stacked as the following tall $N \cdot K$-vector $\mathbf{r}$.

$$\mathbf{r} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_K \end{bmatrix}.$$

The set of all basis vectors at a particular frequency f_k and spatial location (x_p, y_p) can be collected into a matrix $\boldsymbol{\Phi}_{k,p} = \begin{bmatrix} \boldsymbol{\phi}_{k,p,1} & \boldsymbol{\phi}_{k,p,2} & \cdots & \boldsymbol{\phi}_{k,p,M} \end{bmatrix}$. In the same manner, the $\mathbf{b}_m$ vectors can be concatenated into a matrix $\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_M \end{bmatrix}$. These two matrices are related by the expression $\boldsymbol{\Phi}_{k,p} = \mathbf{B} \bullet (\boldsymbol{\varepsilon}_{k,p} \mathbf{1}_M^T)$, where $\mathbf{M}_1 \bullet \mathbf{M}_2$ is the elementwise multiplication of matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, and $\mathbf{1}_M$ is the M-vector of all ones. The factor $\mathbf{B}$ is subject to design in the anisotropy characterization procedure, but $\boldsymbol{\varepsilon}_{k,p}$ is fundamental to the SAR phase history measurements. The choice of $\mathbf{B}$ is discussed in the second half of this section.

Putting together all frequencies and spatial locations, the overall overcomplete basis $\boldsymbol{\Phi}$ is:

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_{1,1} & \boldsymbol{\Phi}_{1,2} & \cdots & \boldsymbol{\Phi}_{1,P} \\ \boldsymbol{\Phi}_{2,1} & \boldsymbol{\Phi}_{2,2} & \cdots & \boldsymbol{\Phi}_{2,P} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi}_{K,1} & \boldsymbol{\Phi}_{K,2} & \cdots & \boldsymbol{\Phi}_{K,P} \end{bmatrix}.$$
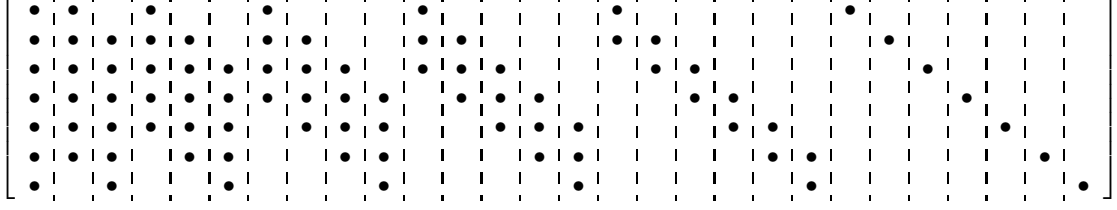
Figure 1. Illustration of matrix $\mathbf{B}$ for $N = 7$.

Each column of the matrix $\mathbf{\Phi}$ is a basis vector of the overcomplete basis.

Defining the length $M{\cdot}P$ vector of coefficients as $\mathbf{a} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,M} & a_{2,1} & \cdots & a_{P,M} \end{bmatrix}^T$, the statement $\mathbf{r} = \mathbf{\Phi a}$ in matrix-vector form is completely equivalent to the summation form (1). In this form, it is readily apparent that there are $N{\cdot}K$ linear equations with $M{\cdot}P$ unknowns and $M > N$. Often in the choice of $\mathbf{B}$, $M \gg N$, so regardless of P and K, $M{\cdot}P > N{\cdot}K$ and the system is an underdetermined set of linear equations. Our formulation also readily deals with the noisy case: $\mathbf{r} = \mathbf{\Phi a} + \mathbf{n}$, where $\mathbf{n}$ is an additive noise term. We delay discussion of solving this inverse problem for $\mathbf{a}$ to Sec. 3.

By forming this $\mathbf{\Phi}$ matrix, the spatial locations are treated jointly within one system of equations, capturing the combined influence of multiple scatterers on individual angle-frequency measurements. The first M elements of $\mathbf{a}$ depend on the first spatial location $p = 1$, elements $M + 1$ to $2M$ of $\mathbf{a}$ depend on the second position $p = 2$, and so on. Thus, by setting up the problem in this manner, it is possible to decompose the phase history data into contributions from different point scatterers at different locations and in the process, characterize amplitude and anisotropy for each one. As a notational convenience, we define $\mathbf{a}_p$ to be the M-vector of coefficients corresponding to position p and:

$$\mathbf{\Phi}_p = \begin{bmatrix} \mathbf{\Phi}_{1,p} \\ \mathbf{\Phi}_{2,p} \\ \vdots \\ \mathbf{\Phi}_{K,p} \end{bmatrix}$$

to be the subset of basis vectors corresponding to position p. There is no requirement that all spatial locations under consideration contain a scatterer. If there is no scatterer at a particular spatial location p, then all of the elements of $\mathbf{a}_p$ should come out to be zero. It is thus possible to use a grid of pixels as the set of potential spatial locations where scatterers might exist. We now discuss the specific choice of basis vectors for the overcomplete basis $\mathbf{\Phi}$.

2.2. Choice of Basis Vectors

The overcomplete basis set is to be chosen such that its cardinality is much greater than the dimension of θ and linear combinations of very few basis vectors accurately represent plausible angle-dependent scattering behaviors. In the selection of the overcomplete basis $\mathbf{\Phi}$, we are free to choose $\mathbf{B}$; the choice of $\mathbf{B}$ is a way to incorporate prior information about angle-dependent scattering.

Methods employing subaperture analysis and parametric models expect to find contiguous intervals in θ for which there is non-zero scattering. Similarly here, basis vectors are chosen such that contiguous segments of anisotropy are represented by a single basis vector. However, our formulation allows the representation of non-contiguous segments through the combination of multiple basis vectors. The $\mathbf{b}_m$ are chosen to be pulses with all possible angular extents and all possible starting angles, in other words all widths and shifts. For example, if $N = 7$ and the pulse shape is rectangular, then $\mathbf{b}_1$, the isotropic vector, is $[1\ 1\ 1\ 1\ 1\ 1\ 1]^T$, $\mathbf{b}_2 = [1\ 1\ 1\ 1\ 1\ 1\ 0]^T$, $\mathbf{b}_3 = [0\ 1\ 1\ 1\ 1\ 1\ 1]^T$, and the final pulse with the finest anisotropy $\mathbf{b}_M = [0\ 0\ 0\ 0\ 0\ 0\ 1]^T$. The $\mathbf{b}_m$ have unit maximum amplitude; solving the inverse problem gives the complex amplitude coefficients $\mathbf{a}$. The full set $\mathbf{B}$ for $N = 7$ is illustrated in Fig. 1. The dots represent entries that have a non-zero value and spaces without dots represent zero-valued elements. For this choice of basis vectors, $M = \frac{1}{2}N^2 + \frac{1}{2}N$. Various pulse shapes, not just
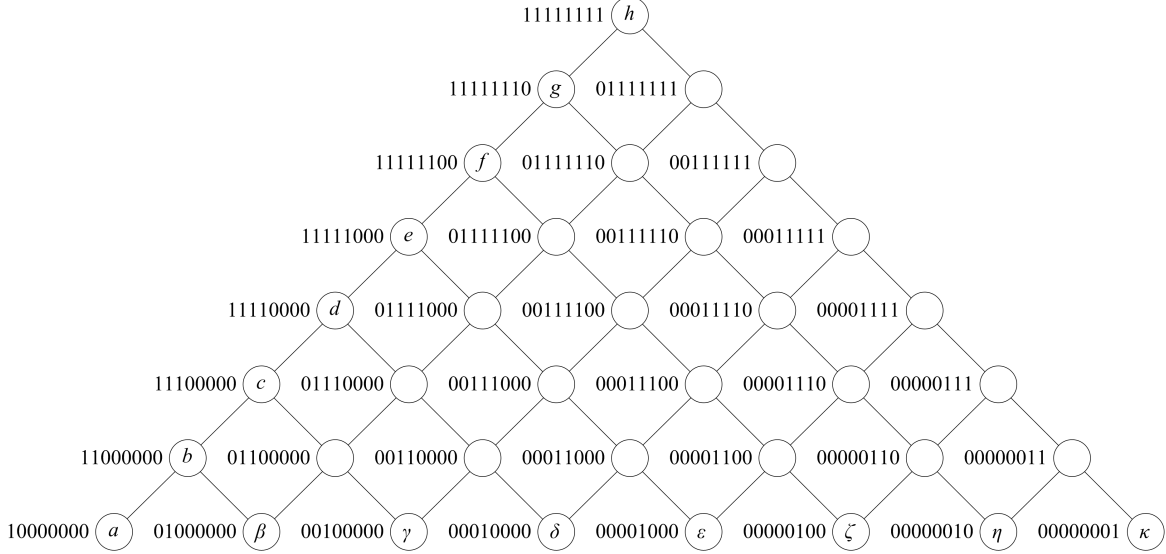
Figure 2. Graph-structured representation of **B**.

rectangular pulses, may be seamlessly enlisted in the overcomplete basis, e.g. triangle, raised triangle, windowed Gaussian, and Hamming pulse shapes.

The collection **B** described above has a nice, intuitive graph-structured interpretation. The vectors $\mathbf{b}_m$ can be arranged as nodes in a graph. The graph is given in Fig. 2 for $N = 8$, with nodes labeled to the left with their corresponding **b** vectors when the pulse shape is rectangular. The labels inside the nodes may be ignored for now. The graph has N levels, with the root node being the isotropic basis vector; traversing down the graph corresponds to decreasing angular extent of anisotropy. A graph of this form is referred to as an N-level *basis graph* in the remainder of this paper. This structure will be useful in the development of greedy algorithms in the next section, which discusses methods to solve the inverse problem $\mathbf{r} = \mathbf{\Phi a}$.

3. METHODS OF SOLUTION TO THE INVERSE PROBLEM

From linear algebra, we know that the underdetermined system of linear equations (1) has no unique solution. Furthermore, the overcomplete basis $\mathbf{\Phi}$ is designed to allow for the representation of the phase history measurements with few basis vectors. Thus the inverse problem is a sparse signal representation problem – among the infinite number of solutions, our formulation favors those solutions that are sparse, i.e. those solutions $\mathbf{a}$ whose ℓ_0-norm is small.

Finding the solution that minimizes $\|\mathbf{a}\|_0$ is a combinatorial optimization problem, but greedy approaches such as matching pursuit, and relaxations such as the ℓ_1 relaxation matching pursuit have been developed. A sparsifying regularization approach incorporating a quasi-Newton optimization algorithm, originally developed for feature-enhanced SAR image formation[14] but applicable to a variety of sparse signal representation problems, is an alternative method. The objective is to find the solution that minimizes the cost function $J(\mathbf{a})$, containing two terms, a data fidelity term and a sparsifying term. Specifically, the form of the cost function is:

$$J(\mathbf{a}) = \|\mathbf{r} - \mathbf{\Phi a}\|_2^2 + \alpha \|\mathbf{a}\|_k^k, \ k < 1. \tag{2}$$

The ℓ_k-norm with $k < 1$ has a sparsifying effect. The scalar α is a regularization parameter that trades off data fidelity and sparsity. Details of this robust method may be found in Ref. 14.

In theory, there is no restriction on the size of the problem that the quasi-Newton method can be applied to. However, the number of columns of $\mathbf{\Phi}$, which is $\mathcal{O}(N^2 P)$ for the overcomplete basis choice discussed in Sec. 2.2,

is restrictive in terms of memory as well as computation for realistic imaging scenarios with hundreds of angle samples and spatial locations. In this section we develop a greedy algorithm with reduced memory requirements taking advantage of the graph structure described in Sec. 2.2.

In the N-level basis graph, the nodes represent the basis vectors in the overcomplete basis. The basis is designed such that a few or often just one basis vector per position p' is sufficient to represent the aspect-dependent scattering function $s(x_{p'}, y_{p'}, \theta)$. Thus, the sparse signal representation problem may be reformulated as a search for a node or a few nodes on the basis graph. In addition to finding nodes, complex amplitudes must also be determined. In general, there are $P > 1$ spatial locations in the problem, and consequently P coexisting basis graphs. Thus, to solve the problem, there is not just one search to be done, but P simultaneous searches. To be most effective, these searches should not be performed independently, but rather should interact and influence each other.

We propose a search strategy akin to guided depth-first search per basis graph, which follows a single path down from the root looking for the goal. Each step in the search is based on a heuristic. If the bottom of the graph is reached without finding the goal, then there is back-tracking also based on the heuristic. Nodes in the basis graph have two children; when progressing downwards during the search, the heuristic is used to determine whether the next step is the left or right child node. Unlike standard graph search problems, in our problem it is not obvious when to terminate the search, so we also need to specify stopping conditions.

Our search heuristic and stopping criterion is founded on solving the inverse problem not with the full set of basis vectors, but with a subset of basis vectors. Let us consider an m-level basis graph, $m \ll N$, with its root at the current node of the search, termed the *guiding graph*, as this subset of basis vectors. The search process will move the guiding graph around through the N-level basis graph.

Intuition about the problem suggests that if the basis vector corresponding to true scattering behavior is not included in the guiding graph when the inverse problem is solved in a sparsity enforcing manner, then the resulting solution coefficient vector $\mathbf{a}$ will have a non-zero coefficient for the basis vector most 'similar' to the truth. In terms of the N-level basis graph, intuition suggests that if the true coefficient is far down in the basis graph, but the inverse problem is solved with only basis vectors from a guiding graph near the top of the N-level basis graph, then coefficients in the first $m - 1$ levels will be zero and coefficients in level m may be non-zero. In the same vein, if the guiding graph is rooted below the true coefficient, then the root coefficient may be non-zero and the coefficients in levels two through m will be zero. Again, intuition suggests that if the guiding graph is such that it contains the true coefficient, then the true coefficient will be non-zero and the rest of the coefficients zero.

Before arriving at the search procedure and heuristics, let us first confirm the above intuition through experimentation for $N = 400$, $m = 8$, and implicitly $P = 1$. The 400 angle samples are over the interval $[-55°, +55°]$, the number of frequencies $K = 3$ with values 7.047 GHz, 7.059 GHz, and 7.070 GHz, and the regularization parameter $\alpha = 150$. The 8-level guiding graph contains 36 nodes. In the first experiment, with results in Fig. 3, the guiding graph is fixed with root at the left-most node of level 200 in the basis graph. The true scattering behavior is varied from isotropic, to anisotropic with medium angular extent, to anisotropic with just one angle sample non-zero. In terms of the 400-level basis graph, the true coefficient is varied, starting at the root node, through all nodes along the left edge of the graph, to the left-most node of level 400, as diagrammed in the left portion of Fig. 3. The large triangle is the 400-level basis graph, the tiny filled triangle is the fixed guiding graph, and the arrows along the left edge indicate the variation of the true node. In the two plots, the angular extent of the true scattering behavior is plotted on the horizontal axis. In the top plot, the coefficient magnitudes for all 36 coefficients associated with the basis vectors in the guiding graph are plotted on the vertical axis, whereas in the bottom plot, coefficient magnitudes are indicated by shading (white is zero) and each horizontal strip is for each of the 36 different coefficients. The coefficient values are obtained by solving the inverse problem using the quasi-Newton method. Most coefficients are zero for all true scattering behaviors in this experiment. Lines on the plots are labeled in correspondence with node labels in Fig. 2. The figure shows that in agreement with intuition, in the regime where the guiding graph is below the true coefficient, the root node (node h) is non-zero. In the regime where the guiding graph covers the true coefficient, the correct node is non-zero. Also in agreement with intuition, when the guiding graph is above the true coefficient, the node in the last level closest to the truth (node a) is non-zero and others are zero.
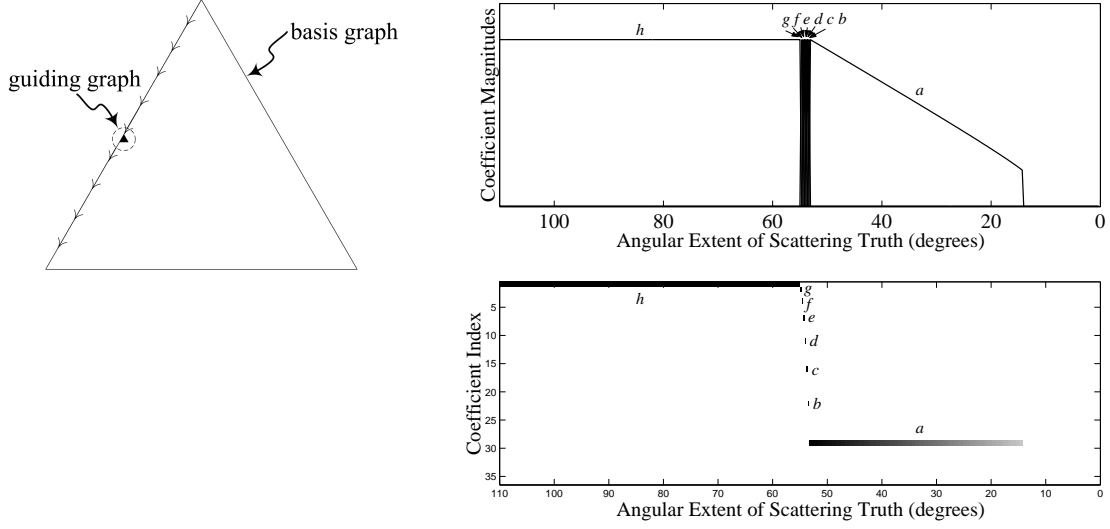
Figure 3. Coefficient magnitudes in m-level guiding graph as true scattering behavior is varied from isotropic to highly anisotropic. The m-level guiding graph is fixed with top node having angular extent $55.3°$ and m^{th} row nodes having angular extent $53.1°$.
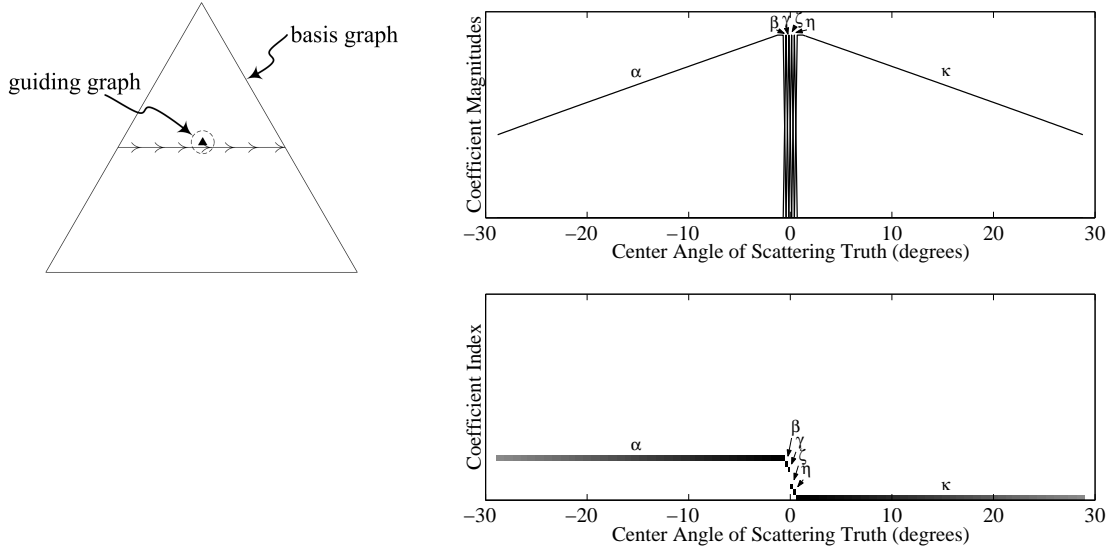

Figure 4. Coefficient magnitudes in m-level guiding graph as center angle of true scattering behavior is varied. The m-level guiding graph is fixed, covering center angles $[-1.0°, +1.0°]$.

The experiment yielding the results of Fig. 4 has the same setup, but the guiding graph is fixed with root at the center node of level 200 instead of the left-most node. The true node is varied from left to right across the basis graph at level 210, three levels below the bottom of the guiding graph, effectively changing the center angle of the anisotropy, but leaving the extent constant. This figure is organized in the same manner as Fig. 3, but the horizontal axis features the center angle rather than angular extent. From these results, first it is apparent that only nodes in the last level of the guiding graph are non-zero, reconfirming results from the previous experiment. Second, it can be seen that when the truth is to the left of the guiding graph, the left-most node of the mth level (node α) is non-zero. Similarly, when the truth is to the right, the right node (node κ) is non-zero; when the

39

truth is underneath the 8-level graph, nodes along the last level (nodes β-η) are non-zero.

Intuition along with these experimental validations suggests simple stopping conditions and heuristics. One stopping criterion is to stop when all of the nodes in level m (nodes α-κ) are zero during the search. A heuristic for the search is also apparent based on the coefficient values of the m nodes in level m. Due to the structure of the basis graph, each node has two children, so the heuristic will be used to determine whether the next guiding graph root will be the left child or the right child of the current guiding graph root node. Based on the second experiment, one reasonable idea is to take the weighted average of the coefficient magnitudes of the bottom level — the search can then be guided towards the side the average indicates to be stronger. The basis used in calculating the given heuristic and stopping criterion has $\mathcal{O}(1)$ columns for each spatial location and $\mathcal{O}(P)$ columns for P spatial locations, providing savings in terms of memory.

For the case of a single spatial location, $P = 1$, the algorithm is as follows. The inverse problem $\mathbf{r} = \mathbf{\Phi}^{(i)}\hat{\mathbf{a}}^{(i)}$ is solved for each iteration i of the search. Then, $\hat{\mathbf{a}}^{(i)}$ is tested for the stopping condition. If the search is to continue, the heuristic is calculated to determine which one of two choices $\mathbf{\Phi}^{(i+1)}$ will be. The initial set of basis vectors $\mathbf{\Phi}^{(1)}$ is the set with the largest angular extent located in the top m levels of the N-level basis graph. For the general case of multiple spatial locations, P searches are performed simultaneously, but not independently. As in the single position case, $\mathbf{r} = \mathbf{\Phi}^{(i)}\hat{\mathbf{a}}^{(i)}$ is still solved on each iteration, but now individual block matrices $\mathbf{\Phi}_p^{(i)}$ evolve based on their corresponding $\hat{\mathbf{a}}_p^{(i)}$. For example, the first spatial location's coefficients $\hat{\mathbf{a}}_1^{(i)}$ may satisfy the stopping condition, in which case $\mathbf{\Phi}_1^{(i+1)} = \mathbf{\Phi}_1^{(i)}$. The second spatial location's coefficients may indicate through the heuristic that the search should proceed to the left child, so $\mathbf{\Phi}_2^{(i)}$ is updated accordingly, and so on. The overall search terminates when all of the $\hat{\mathbf{a}}_p^{(i)}$ satisfy the stopping criterion. The P searches are coupled because the inverse problem is solved jointly for all spatial locations on every iteration. When there are multiple spatial locations, contributions from different positions interact. As stated, the algorithm allows for contributions from more than one basis vector per spatial location in the final solution, but those basis vectors must be within the span of a guiding graph. The guiding graph may be enlarged to allow for contributions from disparate basis vectors at additional expense, the extreme being to take the guiding graph as the full basis graph.

A number of variations to the basic algorithm presented above may be made that further reduce memory or computation. First, the back-tracking component of the algorithm may be removed; if the search terminates before reaching a leaf, then this does not change anything. Without back-tracking, the search becomes greedier and takes $\mathcal{O}(N)$ iterations, whereas with back-tracking there are $\mathcal{O}(N^2)$ iterations. The guiding graph need not be an m-level basis graph; for example, the graph may be thinned and include the top level, bottom level, and a few intermediate levels rather than all intermediate levels. A further approximation can be introduced into the search without back-tracking to reduce the average-case dependence of the number of total basis vectors on P. We can fix the contribution from a spatial location after its coefficients have been found. In the algorithm, this implies that once the stopping criterion is met and maintained for a few iterations at position p, the observation data $\mathbf{r}$ is updated to be $\mathbf{r}' = \mathbf{r} - \mathbf{\Phi}_p\mathbf{a}_p$, and $\mathbf{\Phi}_p$ is removed from matrix $\mathbf{\Phi}$, thereby reducing the number of columns in $\mathbf{\Phi}$. This list of variations is far from exhaustive. The next section gives examples of using our formulation for anisotropy characterization.

4. EXAMPLES

In this section, we present three examples of anisotropy characterization; the first uses the quasi-Newton method on a scene with XPatch data of canonical point scattering, the second uses the graph-structured algorithm on synthetic data, and the last uses the graph-structured algorithm on the backhoe dataset. The first two examples are mainly for illustrative purposes.

In this first example, there are four spatial locations (pixels) at $(0, 0)$, $(0, \frac{1}{2})$, $(\frac{1}{2}, 0)$, and $(\frac{1}{2}, \frac{1}{2})$ meters. We use measurements at $K = 3$ frequencies 9 GHz, 9.016 GHz, and 9.032 GHz over the $N = 50$ angle samples equally spaced over a 98° aperture. Illustrating the fact that all spatial locations need not contain point scattering centers, in this example, two of the spatial locations have no scatterers. The scattering centers at the other two spatial locations exhibit realistic, i.e. from XPatch predictions, aspect-dependent scattering behavior. We solve $\mathbf{r} = \mathbf{\Phi}\mathbf{a}$ using the pseudo-inverse to obtain the least-squares solution as a baseline for comparison. We also use
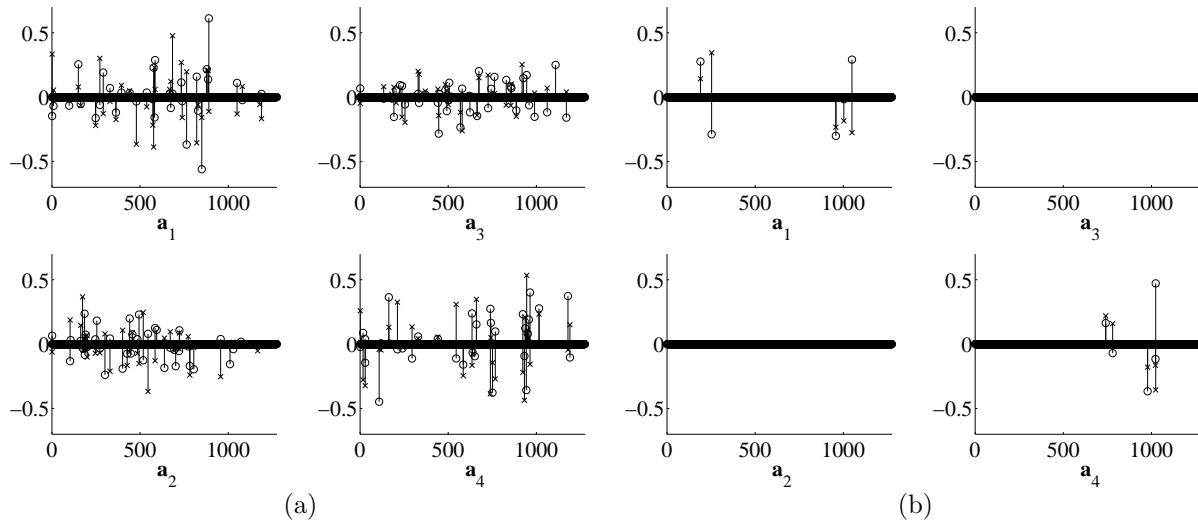
Figure 5. Coefficient vectors **a** from (a) least-squares solution and (b) sparse solution, with real part ○ and imaginary part ×.

the quasi-Newton method with the ℓ_k-norm having $k = 0.1$ and the regularization parameter $\alpha = 1$ to obtain a solution.

Fig. 5 gives stem plots that show the values of the coefficients in the solution **a** vectors, the least-squares solution on the left and the solution with regularization on the right. The stems topped by ○ give the value of the real part and the stems topped by ×, the imaginary part. For $N = 50$, there are $M = 1275$ basis vectors per spatial location; the subplots enumerate the corresponding 1275 coefficients from left to right as in Fig. 1. As expected, the **a** vector on the right is much sparser due to the $\ell_{0.1}$ regularization term. In fact, the coefficients corresponding to spatial locations without scattering centers are all nearly zero in the sparse solution, whereas the least-squares solution has many large-valued coefficients.

Now let us inspect what these coefficients map to in terms of estimated $s(x_p, y_p, \theta)$ functions. Fig. 6 shows the magnitude of the solutions in blue overlaid on the underlying truth in black. The sparse solution is more accurate in its representation of the underlying truth than the least-squares solution, a consequence of the fact that basis vectors of contiguous anisotropy are fairly good at sparsely representing realistic aspect-dependence. It should be noted that the least-squares solution perfectly matches the measurement vector **r**, whereas the sparse solution does not, but data fit is not our primary concern. It should also be noted that if we were to perform image formation without anisotropy characterization, we would have four likely inaccurate pixel values, rather than four accurate functions of θ.

The first example illustrated the importance of sparsity and gave an indication that contiguous basis vectors are a reasonable choice. The second example, with 7 scattering centers and $N = 1541$ angles, shows the operation of the graph-structured algorithm using synthetic data. The aperture is from $-10°$ to $+100°$ and the scattering centers have anisotropy of varying angular extents with a raised triangle pulse shape. Note that in this example, we use raised triangle pulse shapes for the $\mathbf{b}_m$ vectors as well. As a preprocessing step, we first locate the scattering centers by peak extraction on a conventionally formed image. In this example, the extracted scatterer locations are within 4 mm of the truth, with the main source of error being the discrete grid of pixels in the conventionally-formed image. The conventional image and the extracted scattering center locations are shown in Fig. 7c. Then with $P = 7$ and using measurements at the $K = 3$ frequencies 7.047 GHz, 7.059 GHz, and 7.070 GHz, we run the graph-structured algorithm without back-tracking and with the search heuristic and stopping condition discussed in Sec. 3. The guiding graph has 16 levels.

The simultaneous searches are shown in Fig. 7a, where the 1541-level basis graph is indicated by the triangular outline. The resulting $s(x_p, y_p, \theta)$ estimate magnitudes are shown in Fig. 7b as blue lines overlaid on the black
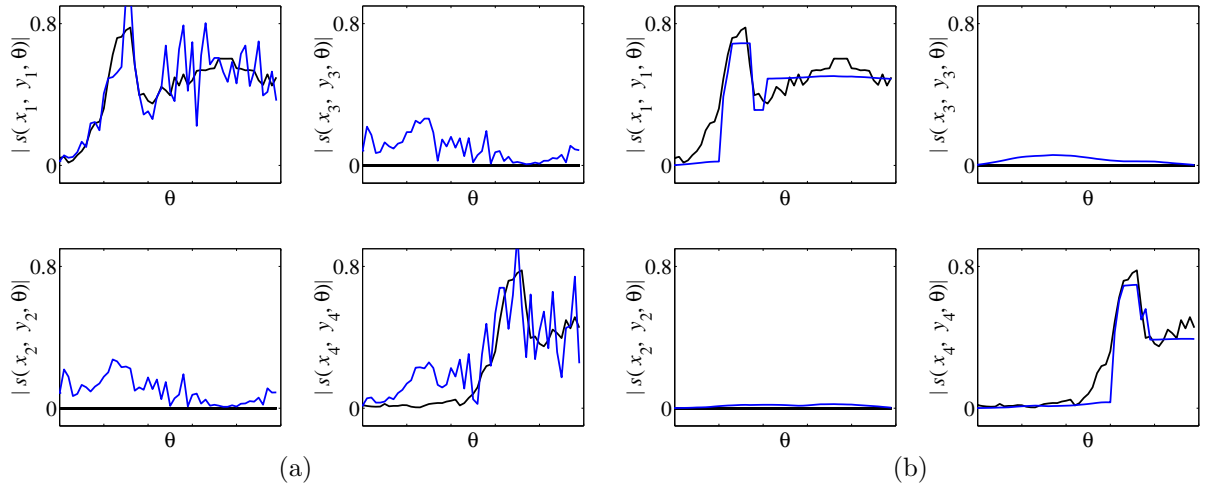
Figure 6. Magnitude of characterized anisotropy from (a) least-squares solution and (b) sparse solution, plotted in blue overlaid on truth.
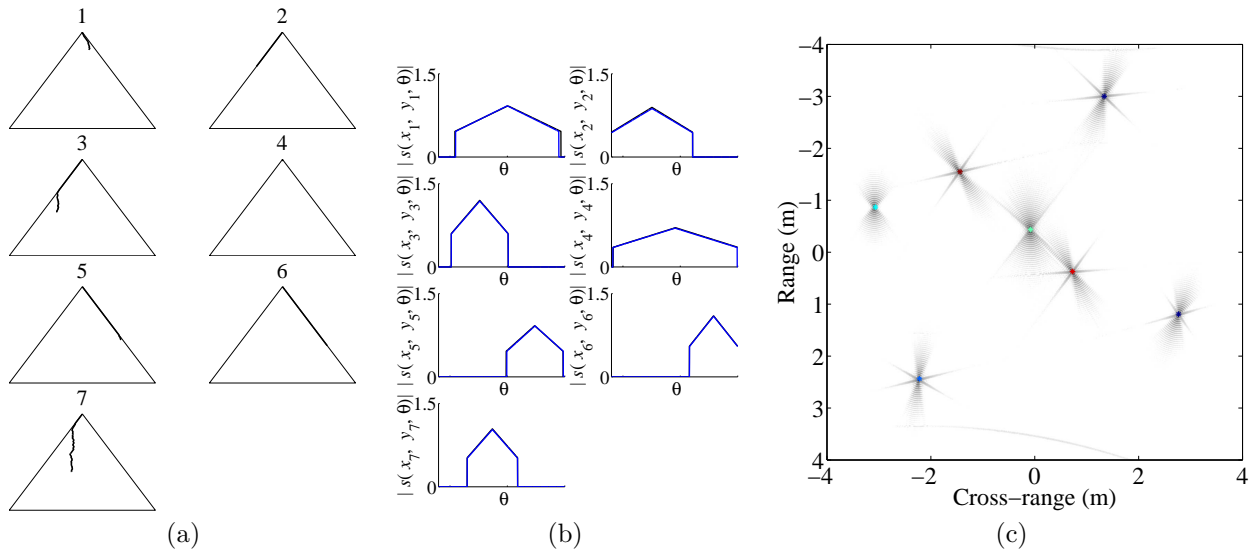

Figure 7. Results of example using graph-structured algorithm: (a) search paths, (b) solution $s(x_p, y_p, \theta)$ overlaid on truth, and (c) color-coded anisotropy center.
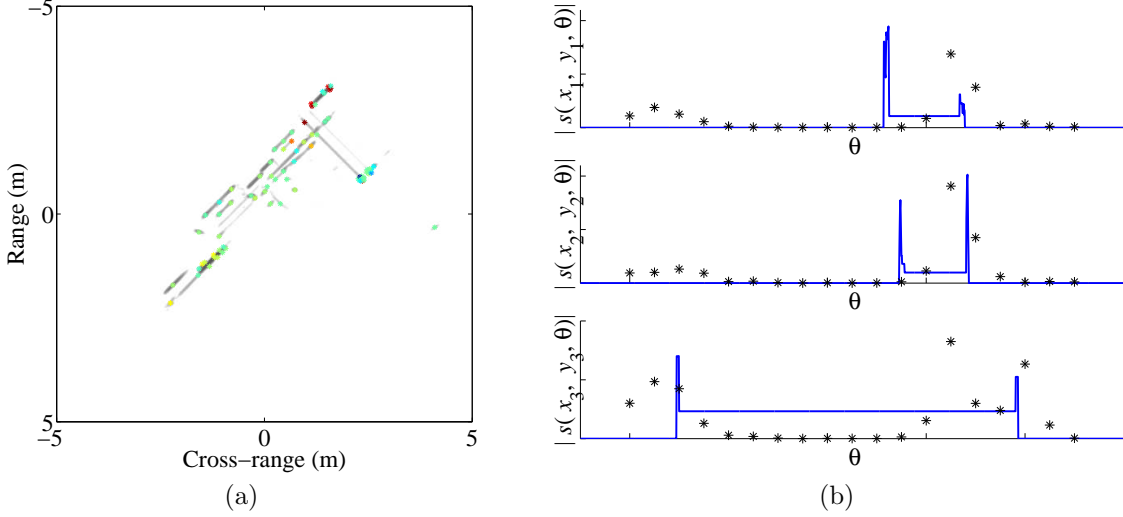
Figure 8. Results of backhoe example using graph-structured algorithm: (a) color-coded anisotropy center, and (b) sample characterized anisotropies.

truth. The estimates are nearly indistinguishable from the true anisotropies. The center of the anisotropy is indicated in Fig. 7c as a color, where dark red corresponds to the center angle closest to $-10°$ and blue corresponds to the center angle closest to $100°$, with the colors cycling red to green to blue. The results are accurate, but moreover, the search paths are fairly direct and would not require back-tracking even if it were available. We see that the formulation is not restricted to rectangle-shaped basis vectors — other pulse shapes may be used as well, as long as they can be used to sparsely represent plausible anisotropy.

The final example, with a dataset from the Backhoe Data Dome,[15] uses the same algorithm as the previous example. The data also has $N = 1541$ angle samples over an aperture from $-10°$ to $+100°$. $P = 75$ spatial locations are extracted from a composite image of conventionally-formed subaperture images[11] and then the greedy graph-structured algorithm is applied to the data with $K = 3$ and frequencies 7.047 GHz, 9.994 GHz, and 12.953 GHz. The solution is displayed in Fig. 8a, also color-coded. The characterized anisotropies of a few scattering centers are also shown in Fig. 8. The solid line indicates our solution, whereas the asterisks show subaperture pixel values from conventional imaging with overlapping $20°$ subapertures.[11] The vertical axes for the line and the set of asterisks are scaled differently to allow comparison. For the first two scattering centers, the subaperture pixel values indicate contiguous extents of anisotropy and our algorithm also detects strong responses at those angles. However, the type of solution we are able to produce is more detailed in θ, especially because the results indicate that anisotropy persistence is not matched to subaperture width. In the third scattering center, the subaperture pixel values indicate two disjoint segments of anisotropy. However, the greedy algorithm may only use basis vectors that lie within a guiding graph to explain the anisotropy. Nevertheless, our algorithm does the best it can to produce two peaks via a positive-valued guiding graph root coefficient and negative-valued leaf coefficient. Multiple candidate search is an approach that would allow for better performance in such instances.

5. CONCLUSION

We have presented a novel approach to SAR image formation. The methodology is general in that it can be applied to a wide variety of overcomplete basis representations. Here we have focused on its utility in describing anisotropic scattering behavior of complex reflections in wide-angle SAR data. The primary advantage of the approach derives from a convenient organization of the basis vectors. The structure allows for a computationally efficient search for the solution of a large sparse regularized inverse problem by evaluating a subset of basis vectors at each iteration. The method demonstrated excellent results on synthetic data, but more importantly, characterized anisotropy to a level of fine detail not possible with subaperture analysis on complex scenes such as

the backhoe dataset. Future work will consider extending the formulation to incorporate another issue that arises in wide-angle imaging, i.e. that certain scattering mechanisms appear to move or migrate in spatial location as a function of aspect angle. Also, we may consider extending the cost function (2) to include preferences other than just sparsity among basis coefficients.

ACKNOWLEDGMENTS

REFERENCES

1. L. C. Potter and R. L. Moses, "Attributed scattering centers for SAR ATR," *IEEE Trans. Image Processing* **6**, pp. 79–91, Jan. 1997.
2. L. C. Trintinalia, R. Bhalla, and H. Ling, "Scattering center parameterization of wide-angle backscattered data using adaptive Gaussian representation," *IEEE Trans. Antennas Propagat.* **45**, pp. 1664–1668, Nov. 1997.
3. M. J. Gerry, L. C. Potter, I. J. Gupta, and A. van der Merwe, "A parametric model for synthetic aperture radar measurements," *IEEE Trans. Antennas Propagat.* **47**, pp. 1179–1188, July 1999.
4. A. Genell and D. Lösaus, "On the Cramer Rao bound in determining scattering center parameters using high resolution radar," Master's thesis, Chalmers University of Technology and Göteborg University, Göteborg, Sweden, 2001.
5. M. R. Allen and L. E. Hoff, "Wide-angle wideband SAR matched filter image formation for enhanced detection performance," in *SPIE Symposium, Algorithms for Synthetic Aperture Radar Imagery*, D. A. Giglio, ed., (Orlando, Florida), Apr. 1994.
6. L. R. Flake, S. C. Ahalt, and A. K. Krishnamurthy, "Detecting anisotropic scattering with hidden Markov models," *IEE Proceedings - Radar, Sonar & Navigation* **144**, pp. 81–86, Apr. 1997.
7. P. Runkle, L. H. Nguyen, J. H. McClellan, and L. Carin, "Multi-aspect target detection for SAR imagery using hidden Markov models," *IEEE Trans. Geosci. Remote Sensing* **39**, pp. 46–55, Jan. 2001.
8. B. Krishnapuram, J. Sichina, and L. Carin, "Physics-based detection of targets in SAR imagery using support vector machines," *IEEE Sensors J.* **3**, pp. 147–157, Apr. 2003.
9. L. Ferro-Famil, A. Reigber, E. Pottier, and W.-M. Boerner, "Scene characterization using subaperture polarimetric SAR data," *IEEE Trans. Geosci. Remote Sensing* **41**, pp. 2264–2276, Oct. 2003.
10. R. L. Moses, L. C. Potter, and M. Çetin, "Wide angle SAR imaging," in *SPIE Defense and Security Symposium, Algorithms for Synthetic Aperture Radar Imagery XI*, E. G. Zelnio and F. D. Garber, eds., (Orlando, Florida), Apr. 2004.
11. M. Çetin and R. L. Moses, "SAR imaging from partial-aperture data with frequency-band omissions," in *SPIE Defense and Security Symposium, Algorithms for Synthetic Aperture Radar Imagery XII*, E. G. Zelnio and F. D. Garber, eds., (Orlando, Florida), Mar. 2005.
12. A. J. Kim, J. W. Fisher, III, and A. S. Willsky, "Detection and analysis of anisotropic scattering in SAR data," *Multidimensional Systems and Signal Processing* **14**, pp. 49–82, Jan. 2003.
13. D. M. Malioutov, "A sparse signal reconstruction perspective for source localization with sensor arrays," Master's thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2003.
14. M. Çetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on non-quadratic regularization," *IEEE Trans. Image Processing* **10**, pp. 623–631, Apr. 1997.
15. "Backhoe data dome and Visual-D challenge problem." Available at Air Force Research Laboratory Sensor Data Management System (https://www.sdms.afrl.af.mil/main.php), 2004.

WIDE-ANGLE SAR IMAGE FORMATION WITH MIGRATORY SCATTERING CENTERS AND REGULARIZATION IN HOUGH SPACE

Kush R. Varshney, Müjdat Çetin, John W. Fisher III, and Alan S. Willsky

Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, Massachusetts 02139
`krv@mit.edu, mcetin@mit.edu, fisher@csail.mit.edu, willsky@mit.edu`

ABSTRACT

Wide-angle synthetic aperture radar imaging presents numerous challenges, but also opportunities to extract object-level information. We present a methodology using an overcomplete dictionary and sparsifying regularization to characterize anisotropy (aspect-dependent scattering amplitude), and migration (aspect-dependent scattering center spatial location), into the image formation process. We also introduce regularization terms in the normal parameter space of the Hough transform that favor solutions with sparsity along a line and consequently parsimony in the representation of glint anisotropy. The characterization of scatterer migration directly gives information about size and shape of objects in the spatial domain and such information can also be inferred from the parsimonious representations we extract for glint-type scattering.

1. INTRODUCTION

The ultimate goal in imaging is understanding what is out in the scene being observed. First steps towards this goal include the collection of measurements and the formation of imagery from those measurements. In synthetic aperture radar (SAR) imaging, data collected over wide-angle apertures permits, in principle, the reconstruction of images with high cross-range resolution. However, conventional SAR image formation techniques, such as the polar format algorithm [1], do not account for certain physical phenomena that arise in wide-angle imaging, leading to inaccurate scattering estimates. In addition, conventional techniques do not extract all possible information from SAR measurements that could be used in higher level scene understanding tasks. In this paper, we propose methods that mitigate these shortcomings of conventional image formation techniques.

In spotlight-mode SAR, measurements are acquired using a radar set mounted on an aircraft. As the aircraft proceeds along its flight path, the radar is continually steered so that it illuminates the same ground patch from all aspect angles of data collection. Recent advances in navigation and avionics technologies now allow long flight paths, or wide-angle apertures. However, dependence of scattering behavior on aspect angle, termed anisotropy, becomes an issue because objects are viewed from different sides rather

than from nearly the same point of view. For example, a mirror or flat metal sheet may reflect strongly when viewed straight on, but barely reflect at all from an oblique angle. This is in opposition to narrow-angle imaging, where it is a fairly reasonable assumption that scattering amplitude is constant over the aperture. In addition, certain scattering mechanisms, such as tophats and cylinders, appear to migrate or move in their spatial location as a function of aspect angle with wide-angle apertures [2].

There are various approaches for anisotropy characterization including parametric methods [3, 4, 5] and methods based on sub-aperture analysis, in which the full collection of SAR measurements is divided into smaller segments covering only parts of the wide-angle aperture and a different image is formed for each sub-aperture [6, 7, 8]. In our previous work, we developed a method for joint image formation and anisotropy characterization based on an overcomplete dictionary and sparsifying regularization [9]. The characterization of migratory scattering has not been given much heed in previous work. In the first part of this paper, we extend our overcomplete dictionary for characterizing anisotropy to account for migratory scattering.

Non-migratory scattering exhibits an interesting relationship between anisotropy and physical extent in the spatial domain. Scattering response over only a very small range of aspect angles, known as glint or flash, arises from long, flat plates, and the thinner the anisotropic response, the longer the spatial extent of the plate. The aspect angle of the glint is also the orientation of the object in space. In the second part of the paper, utilizing Hough transform properties, we introduce new regularization terms to favor solutions that concentrate the representation of glint anisotropy across a spatially distributed area into a single scatterer.

2. SAR OBSERVATION MODEL WITH ANISOTROPIC AND MIGRATORY SCATTERING

The response to radar illumination by the ground patch being observed may be expressed as a complex-valued scattering function $s(x, y)$, where x and y are coordinates on the ground. It is this $s(x, y)$ that conventional image formation techniques attempt to recover. With anisotropy, the scattering function also depends on aspect angle θ, and is thus $s(x, y, \theta)$. At typical operating frequencies of SAR, it is a reasonable assumption that scattering comes from a discrete set of points rather than a continuous field [10]. Measurements are obtained in what is known as the phase history domain. Setting aside migratory scattering in this preliminary exposition, with P point-scatterers the measurements and scattering

function are related by the following expression:

$$r(f,\theta) = \sum_{p=1}^{P} s(x_p, y_p, \theta) e^{-j\frac{4\pi f}{c}(x_p \cos\theta + y_p \sin\theta)}, \quad (1)$$

where c is the speed of propagation and f is frequency. Measurements are discrete, at N angles θ_n and K frequencies f_k.

Another domain in which SAR data may be viewed is the range profile domain. The phase history domain and range profile domain are related by a one-dimensional Fourier transform; ideally, the range profile expression is:

$$\hat{R}(\rho,\theta) = \sum_{p=1}^{P} s(x_p, y_p, \theta)\delta(\rho - x_p\cos\theta - y_p\sin\theta), \quad (2)$$

where ρ parameterizes distance along the line of sight of the radar at aspect angle θ, but because measurements are at a finite set of frequencies f_k within a certain frequency band, there are sidelobe effects. For a single point-scatterer, ideally the range profile is non-zero on a sinusoid $\rho(\theta) = x_0\cos\theta + y_0\sin\theta$.

Now, let us consider migratory scattering. Migration occurs when radar pulses bounce back from the closest surface of a physical object, but the closest surface of the object is different from different viewing angles; the physical object is not really moving, but appears to move in the measurement domain. For the moment restricting ourselves to migration around a circle with center (x_c, y_c) and radius R_0, which could be due to a cylinder or tophat, we note that the point on the circle at angle θ is $(x_c - R_0\cos\theta, y_c - R_0\sin\theta)$. Thus, the sinusoid expression changes to:

$$\rho(\theta) = (x_c - R_0\cos\theta)\cos\theta + (y_c - R_0\sin\theta)\sin\theta$$
$$= x_c\cos\theta + y_c\sin\theta - R_0. \quad (3)$$

Another way to come upon this expression is to consider the fact that at all aspect angles, the surface of the circle is closer to the radar by R_0 than the center. For any general convex shape of migration, the form $x_c\cos\theta + y_c\sin\theta - R(\theta)$ is taken.

In discussing stationary scattering centers, the spatial location (x_p, y_p) and the scattering center p are synonymous. However, care must be taken when discussing migratory scattering centers — some invariant location $(\bar{x}_p, \bar{y}_p)$ is needed to discuss the function $s(\bar{x}_p, \bar{y}_p, \theta)$ for example. We take this invariant spatial location $(\bar{x}_p, \bar{y}_p)$ to be the location the scattering center appears at when $\theta = 0$. When $\theta = 0$, $\bar{x} = x_c - R(0)$ and $\bar{y} = y_c$, leading to the following expression for phase history with migratory point scatterers:

$$r(f,\theta) =$$
$$\sum_{p=1}^{P} s(\bar{x}_p, \bar{y}_p, \theta) e^{-j\frac{4\pi f}{c}\left((\bar{x}_p + R_p(0))\cos\theta + \bar{y}_p\sin\theta - R_p(\theta)\right)}. \quad (4)$$

3. OVERCOMPLETE DICTIONARY AND SPARSIFYING REGULARIZATION FORMULATION

The approach we followed in [9] for anisotropy characterization was to construct an overcomplete expansion of aspect-dependent scattering with $M > N$ atoms per spatial location. We extend that approach here by taking LM atoms per spatial location, where

we do a further expansion in radius of migration with L different values for the radius. (We have once again restricted ourselves to the important case of migration in a circle.)

Specifically, we have PLM coefficients $a_{p,l,m}$ and the overcomplete expansion in the phase history domain is as follows:

$$r(f,\theta) =$$
$$\sum_{p=1}^{P}\sum_{l=1}^{L}\sum_{m=1}^{M} a_{p,l,m} b_m(\theta) e^{-j\frac{4\pi f}{c}\left((\bar{x}_p + R_l)\cos\theta + \bar{y}_p\sin\theta - R_l\right)}.$$
$$(5)$$

The $b_m(\theta)$ represent different persistence widths and center angles of contiguous intervals of anisotropy; more details may be found in [9]. Making the appropriate definitions, the expansion into the overcomplete dictionary can be expressed as:

$$r(f,\theta) = \sum_{p=1}^{P}\sum_{l=1}^{L}\sum_{m=1}^{M} a_{p,l,m}\phi_{p,l,m}(\theta). \quad (6)$$

Each atom $\phi_{p,l,m}(\theta)$ corresponds to a different invariant spatial location, different radius of migration, and different anisotropy. By appropriately stacking the phase history measurements into an $NK \times 1$ vector $\mathbf{r}$, concatenating all of the atoms into an $NK \times LMP$ matrix $\mathbf{\Phi}$, and taking the coefficients as an $LMP \times 1$ vector $\mathbf{a}$, we can also write the overcomplete expansion as $\mathbf{r} = \mathbf{\Phi a}$. The anisotropy and migration characterization problem is thus reduced to solving the inverse problem $\mathbf{r} = \mathbf{\Phi a}$ for the coefficient vector $\mathbf{a}$.

Since $\mathbf{\Phi}$ is overcomplete, we have an underdetermined set of linear equations and the solution is not unique. However, the dictionary is designed such that a sparse collection of atoms approximates commonly encountered scattering behaviors well. Thus, from the infinite subspace of solutions, we favor those solutions $\mathbf{a}$ that are sparse, i.e. having mostly zero coefficients and a few non-zero coefficients, through a sparsifying regularization approach.

The optimally sparse solution is the solution with the minimum ℓ_0-norm, as the ℓ_0-norm simply counts the number of non-zero entries in a vector; however, finding this sparsest solution is a combinatorial optimization problem in general. The approach we take instead is to minimize a regularization cost function of the form:

$$J(\mathbf{a}) = \|\mathbf{r} - \mathbf{\Phi a}\|_2^2 + \alpha\|\mathbf{a}\|_k^k, \ 0 < k < 1, \quad (7)$$

for which efficient optimization techniques exist [11, 9]. The first term is for data fidelity and the second term is for sparsity, with the tradeoff being controlled by the regularization parameter α; we use $k = 0.1$ for the norm in the remainder of this paper.

Let us now consider an example that shows the use of the overcomplete dictionary and sparsifying regularization formulation to characterize both anisotropy and migration within SAR image formation. There is one scattering center in the scene, i.e. $P = 1$, with $N = 15$ angle samples equally-spaced over a $14°$ aperture. The scatterer has a certain anisotropy and has circular migration with radius 0.6 meters. The overcomplete dictionary has $L = 5$ radii, with the R_l being 0, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, and 1 meters. These different R_l are illustrated in Fig. 1 along with the true radius of migration overlaid on an image of the scene formed by conventional processing.

The inverse problem is solved with $K = 5$ frequencies 9.00 GHz, 9.49 GHz, 9.98 GHz, 1.05 GHz, and 1.10 GHz, by the quasi-Newton optimization method of [11]. As a baseline for comparison, we also solve the inverse problem by least-squares, i.e. the
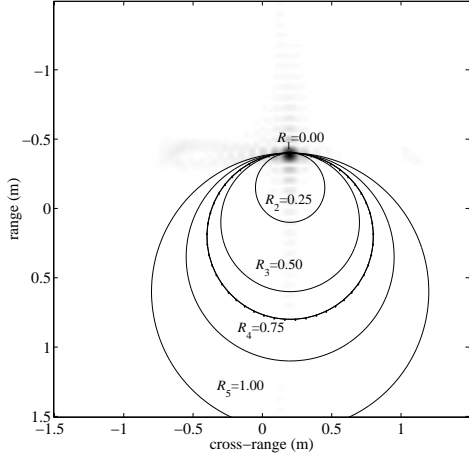
Figure 1: Illustration of atoms of different radii of migration along with true radius of migration, the circle with dots, overlaid on conventionally formed image.
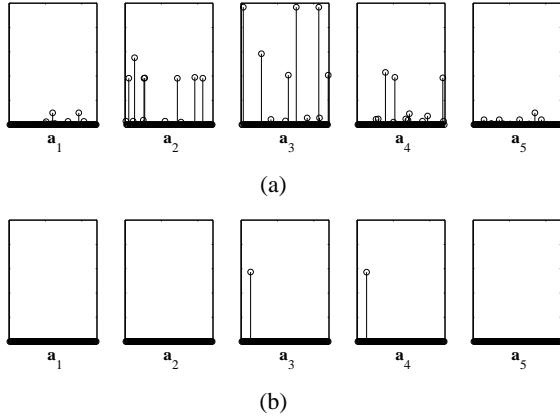

(a)

(b)

Figure 2: Magnitude of coefficients in (a) least-squares solution and (b) sparsifying regularization solution.

regularization parameter $\alpha = 0$ in (7) and we take the minimum norm solution given by $(\mathbf{\Phi}^H\mathbf{\Phi})^{-1}\mathbf{\Phi}^H\mathbf{r}$. The coefficient magnitudes of the two solutions are shown in Fig. 2 as a stem plot; for ease of interpretation, the coefficients $\mathbf{a}_l$ corresponding to each of the $L = 5$ different radii have been put into separate subplots. Within each subplot, the different coefficients correspond to different types of anisotropy; the coefficients on the left correspond to more isotropic scattering and those to the right, to thin, highly anisotropic scattering. In the least-squares solution many coefficients are non-zero in all of the different radii, whereas in the sparse solution, two of the radii, $R_3 = \frac{1}{2}$ and $R_4 = \frac{3}{4}$, have nonzero coefficients corresponding to the true anisotropy. The true radius, 0.6, falls between $\frac{1}{2}$ and $\frac{3}{4}$, so the solution follows expected behavior.

Through the use of atoms in our overcomplete dictionary that correspond to migratory scattering centers, we are able to parsimoniously represent this phenomenon, and consequently model a region in space rather than a single point or pixel because the area covered by the migration is fully described by the atom. The so-

lution compactly represents the scatterer at the object level. Similarly, we would like to find parsimonious representations for stationary scattering centers that cover extended regions in the spatial domain. An approach proposed in the next section uses properties of the Hough Transform.

4. REGULARIZATION IN HOUGH SPACE FOR GLINT ANISOTROPY

Pixels may be treated as scattering centers, but this ignores the fact that a single point scatterer may correspond to a spatially distributed scattering mechanism. One important type of scattering behavior, glint, which comes from long, flat metal plates is nonmigratory, has very thin anisotropy, and corresponds to a line segment in image space oriented at the same angle as the center angle of the anisotropy. A parsimonious representation ought to explain scattering with a single scatterer rather than a collection of scatterers along a line. We extend the regularization cost function (7) to favor sparsity along lines in addition to favoring sparsity among atoms, making use of Hough transform properties and the geometric interpretation they lend.

The Hough transform, which is not a transform in the strict sense, but a method in image analysis for detecting straight lines in binary images [12], uses a ρ-θ normal parameter space that is directly related to the SAR range profile domain, given in expression (2). The normal parameterization uses the angle of a line's normal θ and its algebraic distance ρ from the origin of the image. With x and y as coordinates in the image plane, the equation for a line is $x\cos\theta + y\sin\theta = \rho$.

The parameter space, the ρ-θ plane, and the image space, the x-y plane, are related by the following properties: a point in image space corresponds to a sinusoid in parameter space and a set of points lying on the same line in image space corresponds to a set of sinusoids that intersect at a common point in parameter space. Also, a point in parameter space corresponds to a line in image space and a set of points lying on the same sinusoidal curve in parameter space correspond to a set of lines that intersect at a common point in image space. The Hough transform method of detecting straight lines makes use of these properties.

Let the binary image be such that the background is made up of zero-valued pixels and lines of one-valued pixels. Parameter space is gridded into ρ-θ cells and each one-valued pixel 'votes' for all cells along the sinusoid corresponding to that pixel. If many one-valued pixels are along a common straight line, then their corresponding sinusoids will intersect in one parameter space cell. With parameter space cells acting as accumulators of votes from image domain pixels, a cell with a high count indicates a line in image space. The approach has been extended with different parameters looking for different parameterized curves.

In [13], a Hough space sparsifying regularization approach is employed to enhance and detect straight lines in positive real-valued images by imposing sparsity when taking the image data to the ρ-θ plane. Parameter space cells with small counts are suppressed and cells with large counts are enhanced; thus, non-line features are suppressed and line features are enhanced in image space, making the line detection problem painless. The goals in this paper are different from those in [13] and consequently, the regularization terms are of a different flavor as well: the Hough transform conception of accumulators to detect lines is turned on its head.

The idea is to have sparsity in each cell of the ρ-θ plane rather

than having sparsity among cells. As points on a line in the image domain transform to sinusoids coincident at a point in the range profile domain, sparsity among scatterers in individual ρ-θ cells achieves the goal of sparsity among points on a line. This qualitative description is translated into mathematical terms in the sequel.

The regularization cost $J(\mathbf{a})$ is a function of the coefficient vector $\mathbf{a}$; consequently, in order to work with range profiles, the coefficients must be mapped to that domain first. P separate range profile planes, coming from each of the P scatterers, are required to achieve sparsity among the scatterers in ρ-θ cells.

As mentioned in Sec. 2, the range profile domain and the phase history domain are a single one-dimensional discrete Fourier transform away from each other. Also, the overcomplete dictionary $\mathbf{\Phi}$ is exactly the mapping from coefficients to the phase history domain. However, taking the coefficients through the overcomplete dictionary inherently sums the contributions of each spatial location coherently, which is undesirable when seeking to keep data from the P scatterers separate. Hence, in mapping from coefficients to a set of P range profile planes, a block diagonal matrix $\tilde{\mathbf{\Phi}}$ with $\mathbf{\Phi}_p$, submatrices containing atoms corresponding to spatial location p, on the diagonal is used in conjunction with a matrix $\mathbf{F}$, which is like a DFT matrix. The values are exactly those that would appear in a $K \times K$ DFT matrix, but rearranged to fill an NK by NK area and replicated P times.

Additionally, to select data from a cell $(\rho = \rho_k, \theta = \theta_n)$ in the range profile domain, a matrix $\mathbf{S}_{k,n}$ with P rows and NKP columns composed of mostly zeroes and P ones is used. Specifically, $\mathbf{S}_{k,n}$ is defined as follows with entries indexed by row $i = 1, \ldots, P$, and column $j = 1, \ldots, NKP$:

$$(S_{k,n})_{i,j} = \begin{cases} 1, & j = (k-1)N + n + (i-1)NK \\ 0, & \text{otherwise} \end{cases}. \quad (8)$$

Thus, a length P vector of values for an individual range profile cell (ρ_k, θ_n) is obtained by the multiplication $\mathbf{L}_{k,n}\mathbf{a}$, where $\mathbf{L}_{k,n} = \mathbf{S}_{k,n}\mathbf{F}\tilde{\mathbf{\Phi}}$, and has P rows and MP columns. The $\mathbf{L}_{k,n}$ matrices need not be calculated through matrix multiplication; the $\mathbf{F}\tilde{\mathbf{\Phi}}$ product may be calculated analytically in a straightforward manner based on the discrete Fourier transform of the atoms and the operation $\mathbf{S}_{k,n}$ simply involves extracting out the correct elements from the Fourier transform result.

It follows that for sparsity among scatterers in cell (ρ_k, θ_n), a regularization term of the form $\||\mathbf{L}_{k,n}\mathbf{a}|\|_{0.1}^{0.1}$ is used. Then, continuing to maintain sparsity among atoms, the overall regularization cost function including sparsity in all range profile cells is:

$$J_{\text{line}}(\mathbf{a}) = \|\mathbf{r} - \mathbf{\Phi}\mathbf{a}\|_2^2 + \alpha_0 \|\mathbf{a}\|_{0.1}^{0.1} + \alpha_1 \sum_{k=1}^{K}\sum_{n=1}^{N} \||\mathbf{L}_{k,n}\mathbf{a}|\|_{0.1}^{0.1}, \quad (9)$$

where we have taken the regularization parameters for all cells to be the same. This extended cost function $J_{\text{line}}(\mathbf{a})$ may be minimized using the quasi-Newton method of [11].

We now present an example that uses XPatch data of glint type anisotropy and shows how the extended cost function with both sparsifying terms, the original one and the new one, leads to a parsimonious representation, whereas a cost with either of the sparsifying terms alone with the data fidelity term does not. The scene contains a single scatterer located at $(0, 0)$ with aspect-dependent scattering as shown in Fig. 3. There are $N = 20$ angles over a $19°$ aperture centered around zero degrees. There is a spike in scattering response at $5.5°$, which is the flash or glint. The figure shows
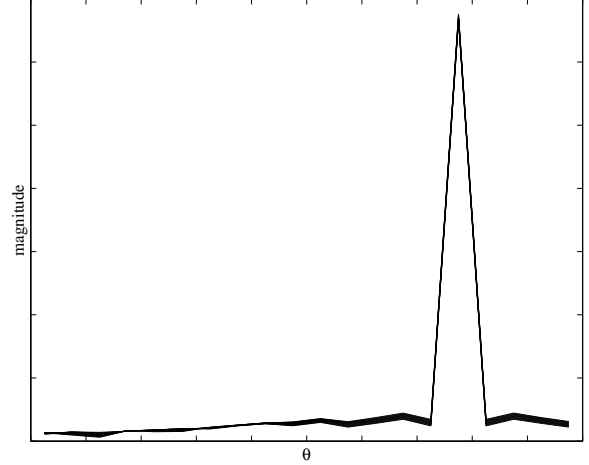

Figure 3: True scattering magnitude of glint anisotropy from XPatch data, with lines for ten different frequency measurements.

the magnitude of the scattering at ten different frequencies in this XPatch data — since there is almost no frequency dependence, the lines are nearly indistinguishable.

In a conventionally formed image using data with a bandwidth of 2 GHz, Fig. 4, the glint shows up as a spread out line segment oriented at $5.5°$. From this image, $P = 24$ pixels are chosen as spatial locations for joint anisotropy characterization and image formation. The spatial locations range from $-\frac{9}{90}$ m to $-\frac{1}{90}$ m in the x direction and from $-\frac{1}{90}$ m to $\frac{2}{90}$ m in the y direction, with a uniform pixel spacing of $\frac{1}{90}$ m in both directions.

Then, with $K = 10$ frequencies in the range 9.00 GHz to 9.14 GHz, the anisotropy is characterized with three different pairs of values for the regularization parameters α_0 and α_1. The first set of regularization parameters is $\alpha_0 = 30$ and $\alpha_1 = 0$, i.e. without the extension to the cost function given in (9). The magnitudes of the coefficients for the twenty-four spatial locations are plotted in Fig. 5, arranged as in an image, and the scattering function magnitudes for each of the spatial locations are given in Fig. 6, also arranged as in an image. The anisotropy has been characterized correctly, but split up and assigned to all of the spatial locations. This solution is parsimonious in atoms per spatial location, but is not parsimonious in the number of spatial locations used.

The second set of regularization parameters is $\alpha_0 = 0$ and $\alpha_1 = 20$: just sparsity among spatial locations along a line. As seen in Fig. 7, the solution in this case has non-zero coefficients at just one spatial location. This spatial location is the closest among all $P = 24$ spatial locations to $(0, 0)$, the true location of the scatterer. However, there are many coefficients with large values, not just one as in the previous case. The coefficients and corresponding atoms are such that they add to match the true anisotropy well, as seen in Fig. 8, but the representation is not parsimonious in terms of atoms per spatial location.

The third set of parameters is chosen such that both sparsifying terms in the regularization cost function are significant. With $\alpha_0 = 30, \alpha_1 = 20$, the solution coefficient vector has only one non-zero coefficient seen in Fig. 9. The coefficient corresponds to an atom with a single non-zero angle sample, shown in Fig. 10, and is thus parsimonious in both spatial locations and atoms.
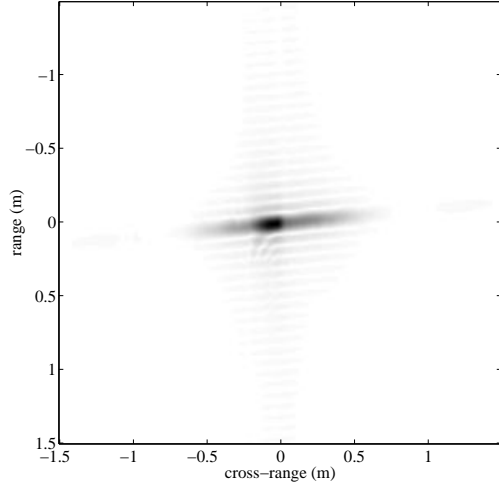
Figure 4: Conventionally formed image of glint anisotropy.

Figure 5: Solution coefficients with $\alpha_0 = 30$, $\alpha_1 = 0$.

Figure 6: Solution scattering magnitudes with $\alpha_0 = 30$, $\alpha_1 = 0$.

Figure 7: Solution coefficients with $\alpha_0 = 0$, $\alpha_1 = 20$.

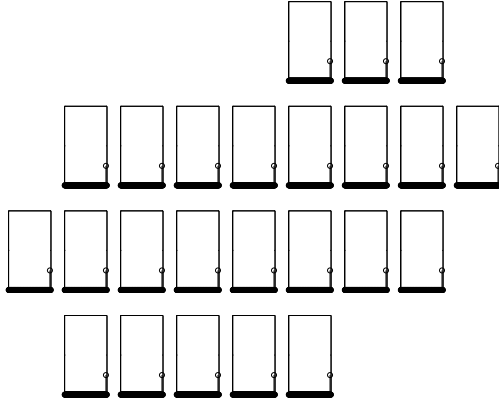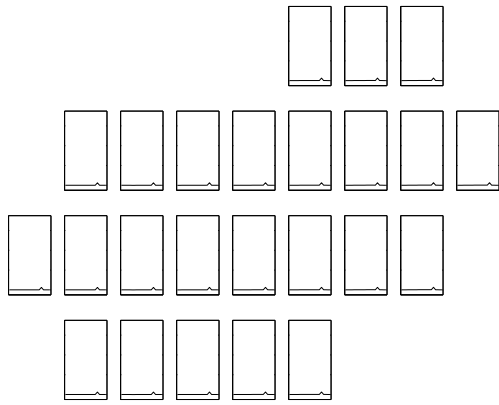Figure 8: Solution scattering magnitudes with $\alpha_0 = 0$, $\alpha_1 = 20$.

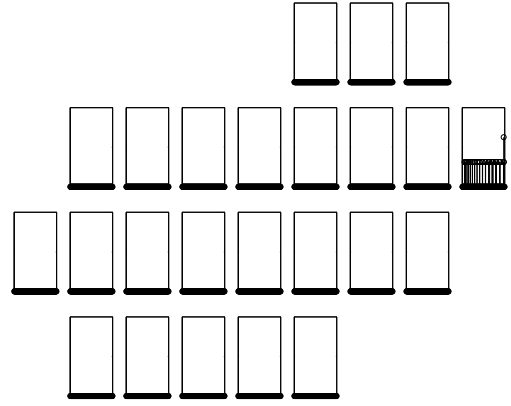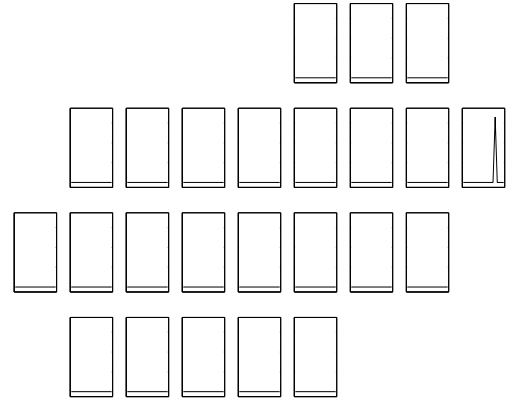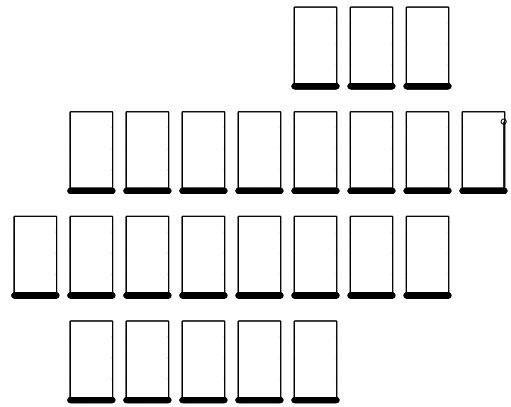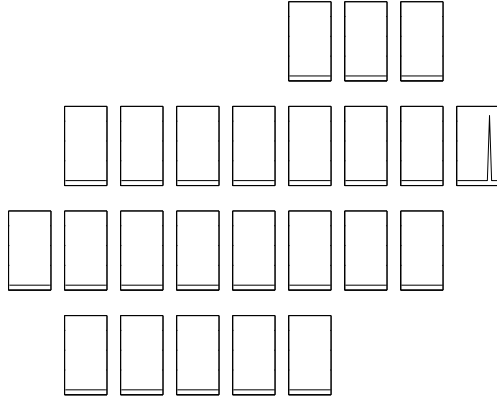Figure 9: Solution coefficients with $\alpha_0 = 30$, $\alpha_1 = 20$.

Figure 10: Solution scattering magnitudes with $\alpha_0 = 30$, $\alpha_1 = 20$.

The original sparsifying regularization cost function has the effect of favoring solutions with sparsity among spatial locations because the vector **a** has coefficients associated with all spatial locations. The additional regularization terms of this section also favor sparsity among spatial locations because spatial locations along a line are general spatial locations as well. However, the distinguishing characteristic of the additional regularization terms is that the favored sparsity is specially adapted for the object-level idea that individual point-scattering centers affect linear regions in space.

Through the example it has been seen that both types of sparsity — sparsity among atoms and sparsity among spatial locations along a line — are necessary in the regularization in order to recover a solution that represents the scattering as coming from a single point and with very thin anisotropy explained by a single atom. With this representation, spatial properties about the object being imaged, such as orientation and physical extent, may be inferred; thin anisotropy corresponds to objects of large physical extent and wider anisotropy to objects with smaller physical extent. Also, the center angle of anisotropy indicates orientation in the spatial domain. Although the same object-level inferences could have been made with the $\alpha_1 = 0$ solution, in that case, P such objects would be indicated rather than one and having P objects all with large spatial extent almost on top of each other does not make physical sense. Points have more meaning than just pixels with aspect-dependent amplitudes.

5. CONCLUSION

We have extended our overcomplete dictionary formulation for anisotropy characterization in SAR imaging to include atoms representing migratory scattering. By doing so, we move beyond standard pixel-based imaging and are able to describe structures with greater semantic meaning within the image formation process. We are also able to find solutions with higher-level meaning in glint-type stationary scattering through an extension to the sparsifying regularization cost function with additional regularization terms operating in Hough space. These object-level descriptions take us a step farther in the scene understanding chain than conventional image formation while also taking into account phenomena such as anisotropy that cause inaccuracies in conventional methods.

As presented, our approach for the characterization of migration limits solutions to migration along a circle, which often arise with tophats and cylinders. The approach can be further extended to handle non-circular migration through the use of subapertures — finding the best circle over a subaperture and then stitching together circular segments over the full wide-angle aperture. Also, glint and sparsity among points on a line is just one imaging scenario, but an important one; other extensions to the regularization cost function for other scattering phenomena and objects may be developed, either based on properties of the Hough normal parameter space or other parameter spaces and domains.

6. REFERENCES

[1] C. V. J. Jakowatz, Jr., D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and P. A. Thompson, *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*. Norwell, Massachusetts: Kluwer Academic Publishers, 1996.

[2] R. Bhalla, A. M. Raynal, H. Ling, J. Moore, and V. J. Velten, "Angular description for 3D scattering centers," in *SPIE Defense and Security Symp., Alg. SAR Imagery XIII*, Apr. 2006.

[3] L. C. Potter and R. L. Moses, "Attributed scattering centers for SAR ATR," *IEEE Trans. Image Processing*, vol. 6, no. 1, pp. 79–91, Jan. 1997.

[4] L. C. Trintinalia, R. Bhalla, and H. Ling, "Scattering center parameterization of wide-angle backscattered data using adaptive Gaussian representation," *IEEE Trans. Antennas Propagat.*, vol. 45, no. 11, pp. 1664–1668, Nov. 1997.

[5] A. Genell and D. Lösaus, "On the Cramer Rao bound in determining scattering center parameters using high resolution radar," Master's thesis, Chalmers University of Technology and Göteborg University, Göteborg, Sweden, 2001.

[6] R. D. Chaney, A. S. Willsky, and L. M. Novak, "Coherent aspect-dependent SAR image formation," in *SPIE Symp., Alg. SAR Imagery*, Apr. 1994.

[7] L. R. Flake, S. C. Ahalt, and A. K. Krishnamurthy, "Detecting anisotropic scattering with hidden Markov models," *IEE P. Radar, Son., Nav.*, vol. 144, no. 2, pp. 81–86, Apr. 1997.

[8] M. Çetin and R. L. Moses, "SAR imaging from partial-aperture data with frequency-band omissions," in *SPIE Defense and Security Symp., Alg. SAR Imagery XII*, Mar. 2005.

[9] K. R. Varshney, M. Çetin, J. W. Fisher, III, and A. S. Willsky, "Joint image formation and anisotropy characterization in wide-angle SAR," in *SPIE Defense and Security Symp., Alg. SAR Imagery XIII*, Apr. 2006.

[10] J. B. Keller, "Geometrical theory of diffraction," *J. Opt. Soc. Amer.*, vol. 52, no. 2, pp. 116–130, Feb. 1962.

[11] M. Çetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Trans. Image Processing*, vol. 10, no. 4, pp. 623–631, Apr. 2001.

[12] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Comm. ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972.

[13] N. Aggarwal and W. C. Karl, "Line detection in images through regularized Hough transform," *IEEE Trans. Image Processing*, vol. 15, no. 3, pp. 582–591, Mar. 2006.

Sparse Representation in Structured Dictionaries with Application to Synthetic Aperture Radar

Kush R. Varshney, *Graduate Student Member, IEEE,* Müjdat Çetin, *Member, IEEE,*
John W. Fisher, III, *Member, IEEE,* and Alan S. Willsky, *Fellow, IEEE*

Abstract— Sparse signal representations and approximations from overcomplete dictionaries have become an invaluable tool recently. In this paper, we develop a new, heuristic, graph-structured, sparse signal representation algorithm for overcomplete dictionaries that can be decomposed into subdictionaries and whose dictionary elements can be arranged in a hierarchy. Around this algorithm, we construct a methodology for advanced image formation in wide-angle synthetic aperture radar (SAR), defining an approach for joint anisotropy characterization and image formation. Additionally, we develop a coordinate descent method for jointly optimizing a parameterized dictionary and recovering a sparse representation using that dictionary. The motivation is to characterize a phenomenon in wide-angle SAR that has not been given much attention before: migratory scattering centers, i.e. scatterers whose apparent spatial location depends on aspect angle. Finally, we address the topic of recovering solutions that are sparse in more than one objective domain by introducing a suitable sparsifying cost function. We encode geometric objectives into SAR image formation through sparsity in two domains, including the normal parameter space of the Hough transform.

Index Terms— sparse signal representations, overcomplete dictionaries, optimization methods, tree searching, inverse problems, synthetic aperture radar, Hough transforms

I. INTRODUCTION

WHETHER for filtering, compression, or higher level tasks such as content understanding, the transformation of signals to domains and representations with desirable properties forms the heart of signal processing. The last decades have seen overcomplete dictionaries and sparse representations take a place in the processing of signals such as those that are multiscale in nature or can be traced to physical phenomena. By sparse, it is explicitly meant that a signal can be adequately represented using a small number of dictionary elements. Sparse signal representation and approximation has proven successful in solving inverse problems arising in a variety of application areas such as array processing [1], time-delay estimation [2], coherent imaging [3], electroencephalography [4], astronomical image restoration [5], and others. Inverse problems may be cast as sparse signal representation or approximation problems in conjunction with dictionaries whose elements have a physical interpretation, having been constructed based on the observation model of a particular application.

Representing a signal $\mathbf{g} \in \mathbb{C}^N$ using an overcomplete dictionary $\{\phi_1, \phi_2, \dots, \phi_M\}$, $M > N$ involves finding coefficients a_m such that $\mathbf{g} = \sum_{m=1}^{M} a_m \phi_m$. Since the dictionary is overcomplete, there is no unique solution for the coefficients; additional constraints or objectives, e.g. sparsity, are needed to specify a unique solution. Among other properties, sparsity and overcomplete dictionaries have been known to deal well with undersampled data, and provide superresolution, parsimony, and robustness to noise. Traditionally, sparsity is measured using the ℓ_0 criterion, which counts the number of non-zero values. The problem of finding the optimally sparse representation, i.e. with minimum $\|\mathbf{a}\|_0^0$ where $\mathbf{a}$ is the set of coefficients taken as a vector in $\mathbb{C}^M$, is a combinatorial optimization problem in general. Due to the difficulty in solving large combinatorial problems, greedy algorithms such as matching pursuit [6] and relaxed formulations such as basis pursuit [7] that are computationally tractable have been developed for general overcomplete dictionaries. Methodologies such as these have been proven to produce optimally sparse solutions under certain conditions on the dictionary [8]–[10]. A sparse signal approximation is a set of coefficients subject to a sparse penalty such that $\|\mathbf{g} - \sum_{m=1}^{M} a_m \phi_m\|_2^2$ is less than a small positive constant.

Oftentimes, the dictionary elements ϕ_m, termed *atoms*, are chosen to have a physical interpretation. Atoms may correspond to different scales, translations, frequencies, and rotations or the dictionary may comprise subdictionaries, often given the name *molecules* [11]. Many popular sparse signal representation methods and algorithms are general and do not exploit natural decompositions of the dictionary into molecules or hierarchical structure that may be present in the collection of atoms. Some approaches do exist in the literature that take advantage of structured dictionaries, e.g. [11]–[16]. A main contribution of this paper is an approximate algorithm for sparse signal representation, related to heuristic search, that uses graphs, one per molecule, constructed with atoms as nodes connected according to hierarchical structure.

In the context of solving inverse problems using sparse

signal representation techniques, the design of atoms based on the observation model is predicated on complete knowledge of the observation process. However, it may be the case that the functional form of the observation process is known, but there is dependence on some parameter or parameters that is not known a priori. In this case, it is of interest to both optimize the dictionary over the unknown parameters and to find sparse solution coefficients. In overcomplete representation contexts other than inverse problems, this can be viewed as signal-dependent dictionary refinement. A second contribution of this work is a coordinate descent approach that simultaneously refines the dictionary and determines a sparse representation.

Notationally, we take $\boldsymbol{\Phi}$ to be a matrix whose columns are atoms from the overcomplete dictionary, and $\boldsymbol{\Phi}(\boldsymbol{\eta})$ to reflect parametric dependence on the set of parameters $\boldsymbol{\eta}$. The matrix for a dictionary with L molecules is the concatenation of L blocks: $[\boldsymbol{\Phi}_1 \cdots \boldsymbol{\Phi}_L]$ or $[\boldsymbol{\Phi}_1(\boldsymbol{\eta}_1) \cdots \boldsymbol{\Phi}_L(\boldsymbol{\eta}_L)]$.

A fundamental premise of sparse signal representation is of underlying sparsity in some domain, but signals may be sparse in more than one complementary, or loosely speaking 'orthogonal,' domain. Accounting for and imposing simultaneous sparsity in multiple domains is important for recovering parsimonious representations. Representational redundancy that may not be apparent in one domain, but apparent in some other domain, can be appropriately reduced through sparsity in that other domain. We consider this problem of sparsity in more than one domain and, as a third contribution, develop a formulation whose objective function includes a carefully composed sparsity term for each domain.

Here we develop a general approach for sparse signal representation or approximation in which we exploit both molecular structure in dictionaries and hierarchical structure within molecules. Additionally, we incorporate dictionary optimization and simultaneously sparsity in multiple domains. While the methods have wider applicability, we focus on modeling wide-angle spotlight-mode synthetic aperture radar (SAR) as an illustrative application. As a consequence, we advance the state of the art in radar imaging as well.

SAR is a technology for producing high quality imagery of the ground using a radar mounted on a moving aircraft. Radar pulses are transmitted and received from many points along the flight path. The full collection of measurements is used to form images; conventional image formation techniques are based on the inverse Fourier transform. In principle, very long flight paths—wide-angle synthetic apertures—which have become possible due to advances in sensor technologies, should allow for the reconstruction of images with high resolution. However, phenomena such as anisotropy and migratory scattering, described in the sequel, which arise in wide-angle imaging scenarios are not accounted for by conventional image formation techniques and cause inaccuracies in reconstructed images. As we proceed in the development of novel sparse signal representation methods for structured dictionaries, we use the methods described herein in a way that does account for such phenomenology.

In Section II we describe a heuristic graph-structured algorithm for producing sparse representations in hierarchical overcomplete dictionaries. Section III expands the scope of the algorithm to dictionaries composed of molecules. The motivating application in Section II and Section III is the characterization of anisotropy in wide-angle SAR measurements, a hurdle that once cleared, not only relieves inaccuracies in image reconstruction, but also provides a wealth of information for understanding and inference tasks such as automatic target recognition. Section IV discusses parameterized dictionaries and the joint optimization of the expansion coefficients and the atoms themselves. The SAR problem investigated in this section is of extracting object-level information as part of the image formation process from migratory scatterers. Section V introduces the objective of sparsity in multiple domains, focusing primarily on the two domain case, specifically with the Hough transform domain and the SAR measurement domain. The applications in Section IV and Section V take steps towards bridging low-level radar signal processing and higher-level object-based processing in ways not seen in the SAR literature before. Section VI provides a summary of our contributions.

II. GRAPH-STRUCTURED ALGORITHM FOR HIERARCHICAL DICTIONARIES

At the outset, we consider a dictionary that does not decompose into molecules and is known and fixed. We look at a particular type of dictionary with a hierarchical arrangement of atoms that permits the construction of a graph with the atoms as nodes. Then, we describe an algorithm based on hill-climbing search, a heuristic search method also known as guided depth-first search. The final part of the section applies the algorithm to the characterization of anisotropy of a point-scattering center from wide-angle SAR measurements.

A. Graph Structure

Oftentimes in overcomplete dictionaries, including for example wavelet packet dictionaries [17], B-spline dictionaries [18], and discrete complex Gabor dictionaries [6], the atoms have a notion of scale and consequently a coarse-scale to fine-scale hierarchy. Translations or rotations are applied at finer scales to create sets of atoms that have a common size but are differentiated in the placement of their region of support; the regions of support may or may not overlap. Some dictionaries are constructed dyadically such that the support of a coarser atom is twice the size of the next finer atom or atoms.

In this work, we consider dictionaries in which the size of the support changes arithmetically rather than geometrically between scales. The matrix $\boldsymbol{\Phi}$ of such a dictionary for one-dimensional signals of length N is illustrated in Fig. 1; the coarsest atom is the first column and the finest atoms are the N right-most columns. A full set of such atoms with all widths and all shifts has large cardinality ($M = \frac{1}{2}N^2 + \frac{1}{2}N$ atoms), but is appealing for inverse problems because of the possibility that a superposition of very few atoms, perhaps just one, corresponds to a physical phenomenon of interest. As discussed in Section II-C, for SAR anisotropy characterization, the signal $\mathbf{g}$ and atoms ϕ_m are such that $\mathbf{g}$ is non-zero for contiguous intervals and zero for other parts of the domain, and is well-represented by few atoms ϕ_m.
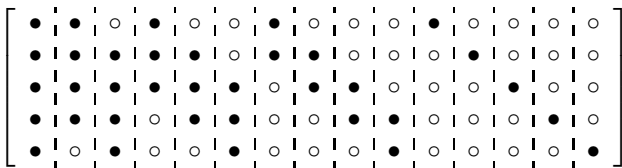
Fig. 1. Illustration of matrix $\mathbf{\Phi}$ for $N = 5$. The solid dots (•) indicate a non-zero value and the empty dots (○) indicate a zero value.
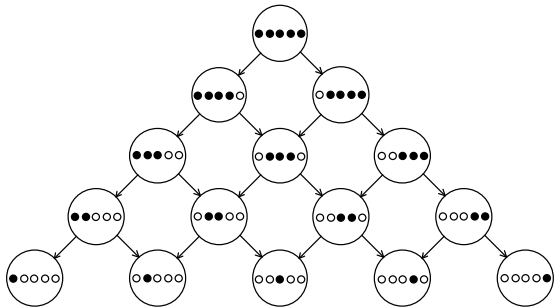

Fig. 2. Illustration of graph structure for overcomplete dictionary, $N = 5$. Coarse-scale atoms are at the top and fine-scale atoms are at the bottom. Different translations are in order from left to right.
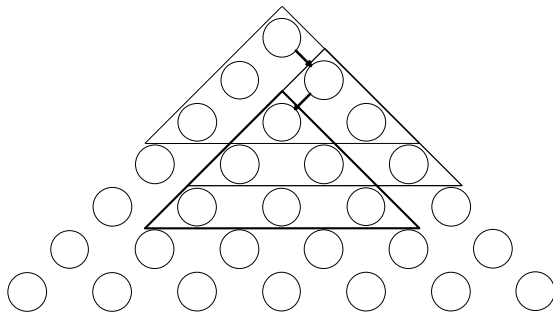

Fig. 3. Illustration of search-based algorithm for $N = 7$, $G = 3$. The guiding graph, a subgraph of the full molecular graph indicated by triangular outline, is moved iteratively to find a sparse representation. The initialization and first two iterations are shown. Molecular graph edges and node labels are omitted.

Due to the regular structure of this type of dictionary, we can take the atoms as nodes and arrange them in a graph. As shown in Fig. 2, the coarsest atom is the root node, the finest atoms are leaves, and the graph has N levels. Each node has two children (except for those at the finest level). It is a weakly connected directed acyclic graph, with a topological sort that is exactly the ordering from left to right of the columns in $\mathbf{\Phi}$ illustrated in Fig. 1. As we proceed, we make use of the graph structure, which we term the *molecular graph*, treating the sparse signal representation problem as a graph search.

B. Algorithm Based on Hill-Climbing

As mentioned in Section I, many general methods for obtaining sparse representations give provably optimal solutions (under certain conditions), but require the same computation and memory regardless of whether the dictionary has structure. As an alternative approach for structured dictionaries, we propose a heuristically-based technique with reduced complexity. The idea to have in mind during the exposition of the algorithm is of a small subgraph, given the name *guiding graph*, iteratively moving through an N-level molecular graph, searching for a parsimonious representation. The specifics of the guiding graph, the search strategy, and search steps are presented below. Fig. 3 illustrates the central idea of the algorithm for a small dictionary; in practice, the dictionary and therefore molecular graph are of much larger cardinality.

We assume that $\mathbf{g}$, the signal to be represented or approximated, can be composed using a few atoms whose nodes are close together in the molecular graph under a common parent node. This assumption is not as restrictive as it may seem: that the signal has a representation with a few atoms is basic for sparsity. Contributing nodes are close together in the graph when the signal is localized in the domain. Prior knowledge can guide the choice of atom shape and standard families of

atoms may be used. The assumptions are reasonable for SAR and other applications that lend themselves to such hierarchical structures.

The problem of finding coefficients $\mathbf{a}$ such that $\mathbf{\Phi a}$ equals or well-approximates $\mathbf{g}$ with few non-zero a_m may be reformulated as a search for a node or a few nodes in the molecular graph. In addition to finding nodes, i.e. atoms ϕ_m that contribute to the expansion, the corresponding coefficient values a_m must also be determined. Numerous search algorithms exist to find nodes in a graph. Blind search algorithms incorporate no prior information to guide the search. In contrast, heuristic search algorithms have some notion of proximity to the goal available during the search process, allowing the search to proceed along paths that are likely to lead to the goal and reduce average-case running time.

Hill-climbing search is an algorithm similar to depth-first search that makes use of a heuristic. In depth-first search, one path is followed from root to leaf in a predetermined way, such as: "always proceed to the left-most unvisited child." In contrast, hill-climbing search will "proceed to the most promising unvisited child based on a heuristic." In both algorithms, if the goal is not found on the way down and the bottom is reached, there is back-tracking. The approach presented here has hill-climbing search as its foundation.

In standard graph search problems, nodes are labeled and the goal of the search is fixed and specified with a label, e.g. "find node K." Thus the stopping criterion for the search is simply whether the label of the current node matches the goal of the search. Also, there is often a notion of intrinsic distance between nodes that leads to simple search heuristics.

When the sparse signal representation problem is reformulated as a search on an N-level molecular graph, stopping criteria and heuristics are not obvious. One clear desideratum is that calculation of both should require less memory and computation than solving the full problem. The guiding graph, chosen to be a G-level molecular graph, $G \ll N$, with its root at the current node of the search, guides the search by providing search heuristics and stopping conditions.

Intuition about the problem suggests that if the atom or atoms that would contribute in an optimally sparse solution are not included in the guiding graph when solving for coefficients

in a sparsity enforcing manner, then the resulting solution will have a non-zero coefficient for the atom most 'similar' to the signal $\mathbf{g}$. In terms of the N-level molecular graph, this suggests that if the optimal sparse representation is far down in the molecular graph, but the problem is solved with a small dictionary containing atoms from a guiding graph near the top of the molecular graph, then coefficients in the first $G-1$ levels will be zero and one or more coefficients in level G non-zero. In the same vein, if the guiding graph is rooted below the optimal representation, then the root coefficient may be non-zero and the coefficients in levels two through G will be zero. If the guiding graph is such that it contains the optimal atoms, then the corresponding coefficients will be non-zero and the rest of the coefficients zero. This intuition is demonstrated empirically; details are in the appendix.

A simple heuristic for the search based on the coefficient values of the G nodes in level G is apparent from the intuition and experimental validation. Due to the structure of the molecular graph, each node has two children, so the heuristic is used to determine whether to proceed to the left child or the right child. We find the center of mass of the bottom level coefficient magnitudes—the search is guided towards the side that contains the center of mass. A stopping criterion is also apparent: stopping when all of the nodes in level G are zero during the search.

Hill-climbing search finds a single node—a single atom. However, the algorithm that we propose is able to find a small subset of atoms due to the guiding graph. When the stopping criterion is met, i.e. when the finest-scale coefficients are all zero in the sparse solution of the representation problem with atoms from the current guiding graph, then that sparse solution is taken as the solution to the full problem. Consequently, the guiding graph allows a subset of atoms rather than a single atom to be used in the representation.

In summary, the algorithm based on the molecular graph and hill-climbing search is as follows.

(1) Initialization: Let $i \leftarrow 1$ and $\mathbf{\Phi}^{(i)} \leftarrow$ atoms from the top G levels of the molecular graph.

(2) Find a sparse $\mathbf{a}^{(i)}$ such that $\mathbf{\Phi}^{(i)}\mathbf{a}^{(i)}$ approximates $\mathbf{g}$.

(3) Calculate weighted sum of bottom row coefficient magnitudes: $\mu \leftarrow \sum_{m=1}^{G} m|a_{\frac{1}{2}G^2 - \frac{1}{2}G + m}^{(i)}|$.

(4) If $\mu = 0$ then stop. Otherwise, $i \leftarrow i+1$. If bottom row nodes are leaves of the molecular graph or both children of the guiding graph have been visited before, then $\mathbf{\Phi}^{(i)} \leftarrow$ atoms from the highest unvisited guiding graph.

Else, $\mathbf{\Phi}^{(i)} \leftarrow$ ($\mu < \frac{G+1}{2}\sum_{m=1}^{G}|a_{\frac{1}{2}G^2 - \frac{1}{2}G + m}^{(i)}|$ and left child unvisited ? atoms from the left child guiding graph : atoms from the right child guiding graph). Iterate to step (2).

The graph-structured algorithm that we propose is able to produce representations in which there are contributions from atoms that lie within the span of a guiding graph. The approximate nature of the approach is controlled by G; by increasing the size of the guiding graph we may, at the expense of increased complexity, draw from a larger subset of atoms in the solution. The smaller problem with $\mathbf{\Phi}^{(i)}\mathbf{a}^{(i)}$ is more tractable than the large problem with $\mathbf{\Phi}\mathbf{a}$.
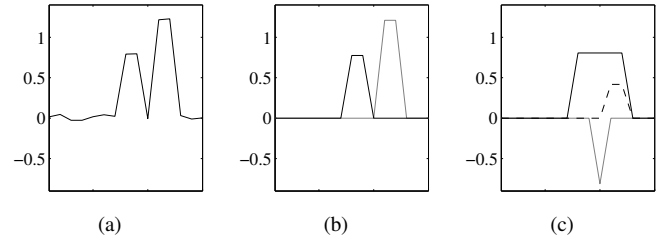

Fig. 4. Comparison of graph-structured algorithm and matching pursuit: (a) the signal $\mathbf{g}$; (b) atoms scaled by coefficients in solution obtained with graph-structured algorithm; (c) atoms scaled by coefficients in solution obtained with matching pursuit.

While any of a number of formulations and techniques may be used to solve the smaller problem, here we use a non-convex, ℓ_p, $p < 1$, relaxation, minimizing the cost function:

$$J\left(\mathbf{a}^{(i)}\right) = \left\|\mathbf{g} - \mathbf{\Phi}^{(i)}\mathbf{a}^{(i)}\right\|_2^2 + \alpha\left\|\mathbf{a}^{(i)}\right\|_p^p, \quad p < 1, \quad (1)$$

by a quasi-Newton technique detailed in [19] to obtain a sparse vector of coefficients $\mathbf{a}^{(i)}$. Each step of the quasi-Newton minimization involves solving a set of M_G linear equations, where M_G is the number of atoms in the guiding graph. Direct solution requires $\mathcal{O}(M_G^3)$ computations. However, the particular matrix involved is Hermitian, positive semidefinite, and usually sparse, so the equations may be solved efficiently via iterative algorithms. We use the conjugate gradient method and terminate it when the residual becomes smaller than a threshold.

The parameter α trades data fidelity, the first term, and sparsity, the second term. The choice of α is important practically and is an open area of research. With α too small, the solution coefficient vector $\mathbf{a}^{(i)}$ is not sparse and the heuristic is not meaningful; the guiding graph strays away from good search paths. With α too large, the algorithm incorrectly terminates early with all zero coefficients in the solution. In this work, we choose the parameter subjectively and can usually set it once for a given problem size. We keep α constant for all iterations of the graph-structured algorithm. Generally, solutions in step (2) of the algorithm are not very sensitive to small perturbations of α. It is possible, however, for a small change in α to cause the number of non-zero elements in the solution to change, but such a change in solution is not necessarily accompanied by a change in the heuristic and stopping criterion. In all examples in this paper, the p of the ℓ_p relaxation is 0.1; for the highly redundant dictionary that is employed, a small value of p results in suitable sparsity.

The search-based procedure we have presented is greedy, but not in the same way as matching pursuit and related algorithms [6], [14]–[16]. A commitment is not made to include an atom in the representation until the final iteration when the stopping criterion is met, and also, atoms within a guiding graph are considered jointly. As the guiding graph slides downwards, any subset of fine-scale atoms can start contributing to the representation. This behavior discourages the assignment of a coarse-scale atom to represent what would be better represented using a few close fine-scale atoms. In some later iteration, a matching pursuit-like algorithm
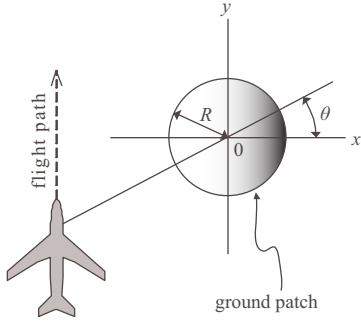
Fig. 5. Ground plane geometry in spotlight-mode SAR.

includes a fine-scale atom with a negative coefficient to cancel extra energy from the coarse-scale atom included earlier. An example of this behavior is given in Fig. 4. For a particular signal **g** and an overcomplete dictionary of boxcar-shaped atoms, solutions are obtained using both the graph-structured algorithm presented in this section and the basic matching pursuit algorithm [6], and compared. Both the graph-structured algorithm and matching pursuit produce solutions that sum to approximate **g**, but the decomposition of the graph-structured algorithm is more atomic.

The algorithm for dictionaries without molecular decomposition is straightforward; its operation in dictionaries with $L > 1$ molecules, which we discuss in Section III, is more interesting. Before reaching that point however, we illustrate the application of this method to anisotropy characterization in SAR.

C. Application to Wide-Angle SAR

Spotlight-mode SAR has an interpretation as a tomographic observation process [20]. As mentioned in Section I, SAR uses a radar mounted on an aircraft to collect measurements. From one point along the aircraft's flight path, the radar transmits a modulated signal in a certain direction, illuminating a portion of the ground known as the ground patch, and receives back scattered energy, which depends on the characteristics of the ground patch. Radar signals are similarly transmitted and received at many points along the flight path. The radar antenna continually changes its look direction to always illuminate the same ground patch. The geometry of data collection in spotlight-mode SAR is illustrated in Fig. 5. Coordinates on the ground plane x, range, and y, cross-range, are centered in the ground patch. Measurements are taken at equally spaced aspect angles θ as the aircraft traverses the flight path. The ground patch, with radius R, is shaded.

The scattering from the ground patch under observation is manifested as an amplitude scaling and phase shift that can be expressed as a complex number at each point. Thus, scattering from the entire ground patch can be characterized by a complex-valued function of two spatial variables $s(x, y)$, which is referred to as the scattering function. Due to the design of the radar signal and the physics of the observation process, the collection of received signals is not $s(x, y)$ directly. Procedures for obtaining $s(x, y)$ from the

measurements are known as image formation. In wide-angle SAR, measurements come from vastly different viewpoints and consequently, scattering behavior shows dependence on θ, referred to as anisotropy, as well as on (x, y) [21]. For example, a mirror-like flat metal sheet reflects strongly when viewed straight on, but barely reflects from an oblique angle. The relationship between the measurements g, obtained over a finite bandwidth of frequencies and over a range of aspect angles, and the anisotropic scattering function $s(x, y, \theta)$ is given by:

$$g(f, \theta) = \iint\limits_{x^2 + y^2 \leq R^2} s(x, y, \theta) e^{-j\frac{4\pi f}{c}(x \cos \theta + y \sin \theta)} dx\, dy, \quad (2)$$

where c is the speed at which electromagnetic radiation propagates. The set of aspect angles θ is inherently discrete, because pulses are transmitted from a discrete set of points along the flight path. The measurements are sampled in frequency f to allow digital processing. The collection of measurements $g(f, \theta)$ is known as the phase history.

The scattering response of objects such as vehicles on the ground is well-approximated by the superposition of responses from point scattering centers when using frequencies and aperture lengths commonly employed in SAR [22]. The anisotropic scattering from a single point-scatterer takes the form $s(x, y, \theta) = s_0(\theta) \cdot \delta(x - x_0, y - y_0)$ and the measurement model is:

$$g(f, \theta) = s_0(\theta) e^{-j\frac{4\pi f}{c}(x_0 \cos \theta + y_0 \sin \theta)}. \quad (3)$$

The phenomenon of anisotropy often manifests as large magnitude scattering in a contiguous interval of θ and small, close to zero magnitude scattering elsewhere. Consequently, the dictionary described in Section II-A containing all widths and all shifts of contiguous intervals is well-suited for obtaining parsimonious representations of anisotropic scattering. An overcomplete expansion is as follows:

$$g(f, \theta) = \sum_{m=1}^{M} a_m b_m(\theta) e^{-j\frac{4\pi f}{c}(x_0 \cos \theta + y_0 \sin \theta)}. \quad (4)$$

Atoms are $\phi_m(\theta) = b_m(\theta) e^{-j\frac{4\pi f}{c}(x_0 \cos \theta + y_0 \sin \theta)}$, where $b_m(\theta)$ are dilations and translations of a common pulse shape. We can use boxcar pulses, Hamming pulses, or other shapes that we expect to encounter. Anisotropy of narrow angular extent comes from physical objects distributed in space and anisotropy of wide angular extent comes from physical objects localized in space; hence the atoms provide a directly meaningful physical interpretation. Appropriately stacking the measurements at different frequencies, we have the sparse signal representation problem with a non-molecular hierarchical dictionary and can obtain solutions using the graph-structured algorithm described above.

D. Anisotropy Characterization of Single Point-Scatterer

We now show anisotropy characterization on SAR phase history measurements from XPatch, a state-of-the-art electromagnetic prediction package, using the graph-structured heuristic method described in this section. A scene containing
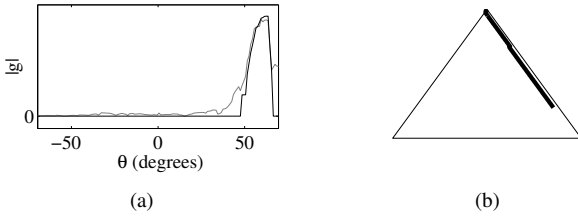
Fig. 6. Single point-scatterer example: (a) aspect-dependent scattering magnitude measurement (gray line) and solution (black line); (b) search path of graph-structured algorithm.

a single scatterer is measured at $N = 140$ aspect angles spaced one degree apart. The scattering magnitude as a function of aspect angle is the gray line plotted in Fig. 6a. (The line shows the measurements at one particular frequency within the frequency band covered by the radar pulse; frequency dependence is minimal and scattering magnitude at all frequencies is nearly the same.)

Using boxcar pulses for atoms in the overcomplete dictionary and a guiding graph of size $G = 32$, we obtain a sparse approximation for the aspect-dependent scattering given by the black line in Fig. 6a. The search path of the graph-structured algorithm is shown in Fig. 6b. The line indicates the location of the root node of the guiding graph within the full molecular graph. When the stopping criterion is met, the atom at the root of the guiding graph is of width 34 samples. The finest atoms that contribute to the approximation have width 4 samples. The sparse solution has 14 non-zero coefficients out of a possible $M = 9870$ coefficients for $N = 140$.

From the solution, it is possible to infer physical properties about the object being imaged because thin anisotropy corresponds to objects of large physical size and wide anisotropy to objects of small physical size. Sparsity and the particular overcomplete dictionary are important because they allow this characterization directly by identifying the coarsest non-zero coefficient.

III. ALGORITHM FOR MOLECULAR DICTIONARIES

In the previous section, we described a search-based algorithm for dictionaries whose atoms have a hierarchy, but did not consider dictionaries that have a molecular decomposition into subdictionaries. In this section, the heuristic algorithm is extended by applying it to dictionaries with $L > 1$ molecules, each individually having a hierarchical structure of atoms. We have L coexisting molecular graphs and thus not just one search, but L simultaneous searches. As we shall see, these searches are not performed independently, but rather interact and influence each other. For joint anisotropy characterization and image formation, the L molecules correspond to L different point-scatterers or spatial locations in the ground patch being imaged.

A. Molecular Dictionaries

Overcomplete dictionaries composed of molecules are fairly common, arising in one of two ways. The first is as the union of two or more orthogonal bases and the second, through dependence on some parameter that takes the same value for one subset of atoms, another value for a subset disjoint from the first, and so on.

An example of the first instance is a dictionary made up of the union of an orthogonal basis of lapped cosines and an orthogonal basis of discrete wavelets that provides atoms to represent tonal and transient components in audio signals [11]; the same idea is used for images as well, taking two different bases together as an overcomplete dictionary, one for periodic textures and one for edges [23]. An example in audio of the second instance is molecules whose atoms share a common fundamental frequency [12]. In the radar imaging example in Section III-D, atoms within molecules share a common (x, y) location and different molecules correspond to different spatial locations.

The two types of decompositions into molecules present different properties. In the first type, different molecules aim to represent very different phenomena and are incoherent from each other, whereas in the second, the molecules correspond to different instances of the same phenomenon and may be highly coherent. In this work, we consider dictionaries whose molecules all have hierarchical structure that permits the construction of molecular graphs, regardless of decomposition type. We use simultaneous searches on all molecular graphs; the difficulty of the problem increases as the coherence between molecules increases.

B. Interacting Searches on Multiple Graphs

The general framework for the graph-structured algorithm with dictionaries containing more than one molecule is the same as for dictionaries without molecules, but with a few key differences. Here the dictionary is of the form $\begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \cdots & \mathbf{\Phi}_L \end{bmatrix}$ with each molecule $\mathbf{\Phi}_l$ having a molecular graph. We assume that all atoms in the dictionary are distinct and that molecules do not share atoms. L guiding graphs iterate through the L molecular graphs, one guiding graph per molecular graph. The vector of coefficients $\mathbf{a}$ also partitions as $\begin{bmatrix} \mathbf{a}_1^T & \mathbf{a}_2^T & \cdots & \mathbf{a}_L^T \end{bmatrix}^T$. L searches are performed simultaneously, as follows.

```
    (1) Initialization: Let i        ←       1 and for
all molecules l      =       1,...,L,  Φ_l^(i)     ← atoms
from the top G levels of molecular graph l.
Φ^(i) ← [ Φ_1^(i)   ···   Φ_L^(i) ].
    (2) Find a sparse a^(i) such that Φ^(i)a^(i)
approximates g.
    (3) For all l      =      1,...,L, calculate weighted
sum of bottom row coefficient magnitudes:
μ_l ← Σ_{m=1}^{G} m|a_{l,½G2-½G+m}^(i)|.
    (4) If Σ_{l=1}^{L} μ_l      =       0 then stop. Otherwise,
i ← i + 1. For all l = 1,...,L, if μ_l = 0, then
Φ_l^(i)      ←       Φ_l^(i-1). Else if bottom row nodes are
leaves of molecular graph l or both children of
guiding graph l have been visited before, then
Φ_l^(i) ← atoms from the highest unvisited guiding
graph. Else,  Φ_l^(i)  ←  (μ_l  <   (G+1)/2 Σ_{m=1}^{G} |a_{l,½G2-½G+m}^(i)|
and left child unvisited ? atoms from the left
child guiding graph : atoms from the right
child guiding graph). Iterate to step (2).
```

Let us emphasize that although the L searches are performed simultaneously, they are not performed independently. The searches are coupled because the inverse problem is solved jointly for all molecules on every iteration; contributions to the reconstruction of $\mathbf{g}$ from all of the molecules interact. There is no notion of molecules when solving the smaller inverse problem $\mathbf{g} \approx \mathbf{\Phi}^{(i)}\mathbf{a}^{(i)}$. The molecular structure only comes into play after $\mathbf{a}^{(i)}$ has been solved, and the heuristics, stopping criteria, and $\mathbf{\Phi}_l^{(i)}$ updates are to be calculated. Since we consider all molecules jointly rather than one at a time as matching pursuit-like algorithms would do, we see similar advantages of the formulation presented here to those seen in Fig. 4 for the single molecule case.

The dictionary used in calculating the heuristic and stopping criterion has $\mathcal{O}(G^2)$ atoms per molecule and $\mathcal{O}(G^2 L)$ atoms for L molecules, instead of $\mathcal{O}(N^2 L)$ atoms used if one were to solve the full inverse problem. However, the graph-structured algorithm requires $\mathcal{O}(N^2)$ iterations, whereas solving the full inverse problem at once requires just one iteration. G is a small constant that is fairly independent of N. For joint anisotropy characterization and image formation, L and N may be in the thousands. The realistic example given in Section III-E would have eighty-nine million atoms if the full problem were solved at once, but the graph-structured approach allows us to only consider a small fraction of them. In the following section, we discuss variations to the algorithm presented thus far that further reduce computation or memory requirements.

C. Algorithmic Variations

The graph-structured algorithm described thus far uses the full hill-climbing search including back-tracking, taking steps of single levels per iteration based on a heuristic employing guiding graphs taking the form of G-level molecular graphs. A number of variations to the basic algorithm may be made; we present a few here, but many others are also possible. Algorithms that use one variation or use a few variations together can be used to solve the sparse signal representation problem. Depending on the size of the problem and the requirements of the application, one algorithm can be selected from this suite of possible algorithms.

1) Hill-climbing without back-tracking: Hill-climbing search always finds the goal node because of back-tracking. In a first variation, we limit the search to disallow back-tracking. This reduces the iterations from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$, but results in a greedier method. If, on a particular example, hill-climbing with back-tracking were to terminate on the first pass down molecular graphs before reaching leaves, then the same operation would be achieved whether the original algorithm or the variation were used. In practice, we often observe termination on the first downward search, including in the example seen in Section II-D and an example presented below in Section III-D.

2) Modified molecular graph: Molecular graphs are structured such that in hill-climbing without back-tracking, one wrong step eliminates many nearby nodes and paths because each node has only two children. The graph may be modified to increase the number of children per node to four for

interior nodes and three for nodes on the edges of the graph, consequently not disallowing as many nodes and paths per search step.

A modified heuristic to go along with this modified graph is to use the G coefficients in level G of the guiding graph as before, but instead of determining whether the center of mass of the coefficient magnitudes is in the left half or the right half, determining which quarter it is in. If the left-most quadrant, then the search proceeds to the node in the next level that is two to the left of the current node. If the middle left quadrant, then the next node is one to the left in the next level, and so on. With these additional edges, search without back-tracking is less greedy with no additional cost, since calculating this modified heuristic is no more costly than calculating the original heuristic.

3) Modified guiding graph and larger steps: The guiding graph need not be a G-level molecular graph; for example, the graph may be thinned and include the top node, nodes in level G, and nodes in a few intermediate levels rather than all intermediate levels, further reducing the number of atoms in $\mathbf{\Phi}^{(i)}$. These atoms are sufficient for calculating the heuristic and stopping condition. Also, searches may take larger steps than moving guiding graphs down just one level per iteration.

4) Removal of stopped molecules: The graph-structured algorithm reduces the number of atoms per molecule from $\mathcal{O}(N^2)$ to $\mathcal{O}(G^2)$, but does nothing to reduce the number of molecules L. A further variation to the hill-climbing search without back-tracking may be introduced that reduces the average-case dependence of the number of atoms on L. It is observed that, despite interactions among contributions from different molecules, once the search on a particular molecule stops it does not restart in general, but may occasionally restart after a few iterations. It is thus natural to consider fixing the contribution from a molecule upon finding its coefficients.

In the algorithm, this implies that once the stopping criterion is met at molecule l, the signal $\mathbf{g}$ is updated to be $\mathbf{g}' = \mathbf{g} - \mathbf{\Phi}_l \mathbf{a}_l$, and $\mathbf{\Phi}_l$ is removed from $\mathbf{\Phi}$, thereby reducing the number of atoms in $\mathbf{\Phi}$. We perform the removal some iterations after the stopping criterion is met and maintained to allow for a possible restart. This variation, though distinct, has some similarity to matching pursuit.

D. Joint Anisotropy Characterization and Image Formation

The problem of joint anisotropy characterization and image formation in wide-angle SAR takes the problem of characterizing anisotropy of a single point-scatterer seen in Section II and extends it to doing so for all points in the ground patch. In other words, whereas standard image formation attempts to recover $s(x, y)$ assuming no dependence on θ, we aim to recover $s(x, y, \theta)$.

The observation model from more than one point is a superposition of terms like (3):

$$g(f, \theta) = \sum_{l=1}^{L} s_l(\theta) e^{-j \frac{4\pi f}{c}(x_l \cos \theta + y_l \sin \theta)}. \quad (5)$$

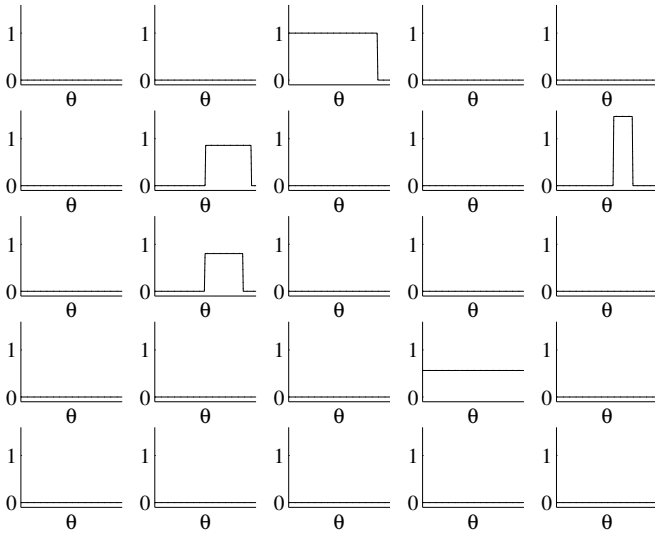The observation model (5) lends itself to an overcomplete

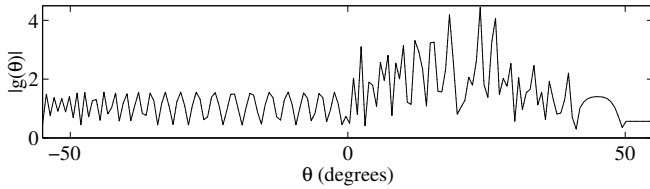Fig. 7.   Scattering magnitude at each spatial location.

Fig. 8.   Phase history measurement magnitude.

expansion of the form:

$$g(f,\theta) = \sum_{l=1}^{L} \sum_{m=1}^{M} a_{lm} b_m(\theta) e^{-j\frac{4\pi f}{c}(x_l \cos\theta + y_l \sin\theta)}, \quad (6)$$

in a similar manner to the single point-scatterer case. Here the dictionary is naturally decomposed into molecules, with each molecule corresponding to a different spatial location (x_l, y_l). We can thus use the methods described above for joint anisotropy characterization and image formation [24].

When performing joint anisotropy characterization and image formation, a grid of pixels in the image to be reconstructed or points of interest identified through preprocessing may be used as the spatial locations (x_l, y_l). We now present an example with $L = 25$ spatial locations in a five by five grid, with rows and columns spaced one meter apart. Unlike Section II-D which uses XPatch data, the synthetic data in this example is matched to the dictionary for illustrative purposes.

This example has $N = 160$ aspect angles equally spaced over a $110°$ aperture. Fig. 7 shows the scattering magnitude at each of the 25 spatial locations arranged as in an image; five of the spatial locations contain boxcar-shaped scattering and the other twenty do not have scatterers. The coherent sum of the scatterers is the phase history measurement $g(f, \theta)$, plotted in Fig. 8 for one frequency.

We recover a signal representation from the phase history measurements using the basic algorithm for molecular dictionaries with guiding graphs of size $G = 8$ and boxcar-shaped atoms. The search paths for the different locations
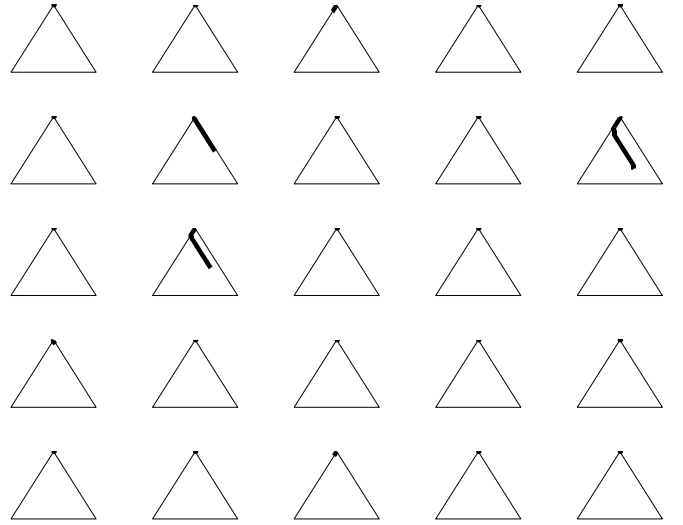

Fig. 9.   Search paths of basic algorithm for molecular dictionaries.

are shown in Fig. 9. The overcomplete dictionary for $N = 160$, $L = 25$ has $322,000$ atoms. In the solution of the sparse signal representation problem, contributions come from exactly the five atoms used to generate the synthetic data; the coefficient values are also recovered. If the solution were to be overlaid on Fig. 7 and Fig. 8, it would not be distinguishable. Looking at the search paths, despite not containing scatterers, a couple of molecules initially iterate nonetheless, but in the end correctly give all zero coefficients. This effect is a result of the interaction between different molecules. The algorithm operates correctly on this synthetic example; a larger example on XPatch data is given below and others may be found in [24], [25].

E. Approaches to Wide-Angle SAR and a Realistic Example

To conclude this section, a large, realistic example with XPatch data is presented. The scene being imaged contains a backhoe-loader, illustrated in Fig. 10a [26]; measurements are taken at $N = 1541$ equally-spaced angles over an aperture ranging from $-10°$ to $100°$. $L = 75$ spatial locations are identified from a composite subaperture image using the method of [27], for which anisotropy is then jointly characterized. The full dictionary for this example has $M = 89,108,325$ atoms. We apply the graph-structured algorithm with all of the variations listed in Section III-C to the problem and obtain seventy-five functions of aspect angle.

The magnitudes of two of these functions are plotted in Fig. 10e and Fig. 10f. In order to provide spatial visualization of the scattering behavior, the magnitude, center angle of anisotropy, and angular extent of anisotropy for each of the spatial locations is indicated by the shading of the markers in Fig. 10b-d.

In the magnitude visualization, light gray is small magnitude and black is high magnitude. Points corresponding to the front bucket of the backhoe-loader have high magnitude. In the visualization of center angle, the left side of the front bucket has responses closer to $-10°$ (light gray) and the
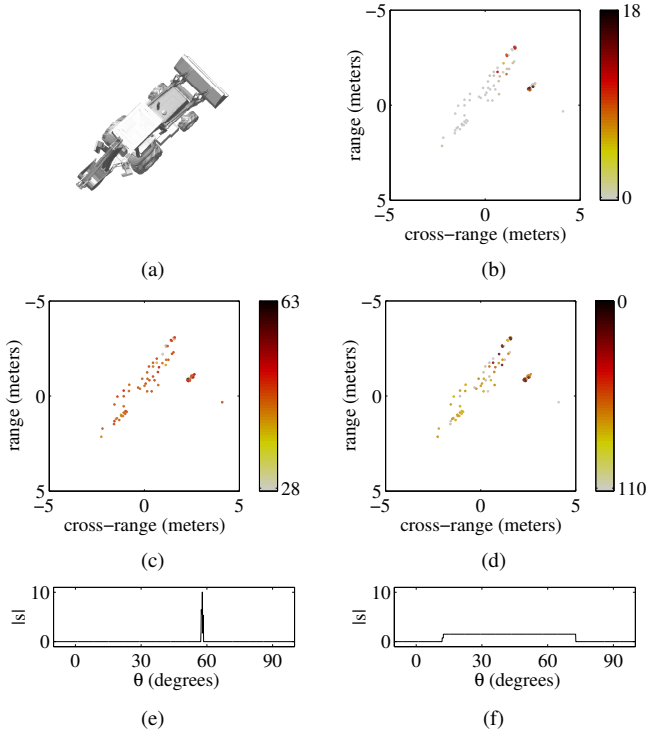
Fig. 10. Backhoe-loader example: (a) illustration of the scene; $L = 75$ spatial locations of interest shaded according to (b) maximum magnitude, (c) center angle of anisotropy (degrees), and (d) angular extent of anisotropy (degrees) in solution; (e)-(f) aspect-dependent scattering solution for two spatial locations.

right side of the front bucket has responses closer to $+100°$ (black). In the angular extent visualization, it can be seen that narrow and wide anisotropy is distributed, but the points on the front bucket with high magnitude also have narrow extent. Overall, one can note from the visualizations that the front bucket flashes on its two sides and the other parts of the backhoe-loader have scattering with smaller magnitude and wider anisotropy.

Through joint anisotropy characterization and image formation, we obtain much more information than a simple image would provide, namely an entire dimension of aspect-dependence. The reflectivities of scatterers with narrow angular persistence, which are lost in Fourier-based image formation, are obtained. The formulation presented here solves for the anisotropy of all spatial locations within one system of equations, taking interactions among scattering centers into account.

The formulation is more flexible than parametric methods for anisotropy characterization such as [28], [29]. Also, solutions have more detail in aspect angle than subaperture methods such as [30]–[33], in which the measurements are divided into smaller segments covering only parts of the wide-angle aperture. Consequently, using the method presented here, angular persistence information can be extracted as in Fig. 10d, which is not possible from subaperture methods. Also, since data from the full wide-angle aperture is used here throughout, cross-range resolution is not reduced as it is with subaperture methods.

IV. DICTIONARY REFINEMENT

In Section II and Section III, the dictionary $\boldsymbol{\Phi}$ is known and fixed, but this need not always be the case. A more ambitious goal is to find the best dictionary under some criteria and an optimally sparse representation jointly. The idea of learning overcomplete dictionaries has been applied in the case that one has many examples of signals $\mathbf{g}$, much more than the number of atoms in $\boldsymbol{\Phi}$, and a dictionary is to be determined that is able to most sparsely represent all of the signals, usually for compression tasks [34], [35]. In inverse problems, where the interest is in extracting physical meaning from the obtained sparse representation for each input signal $\mathbf{g}$, rather than compression of an entire signal class, it is of interest to look at the best dictionary for each input rather than the best dictionary to represent an entire set of training signals. At this point, one could conclude that a dictionary with $\phi_1 = \mathbf{g}$ is optimal and stop. However, we would like to consider dictionaries derived from a parameterized observation model and only consider parameterized atoms, not arbitrary atoms. In this section we propose and demonstrate a formulation for joint optimization to achieve a sparse coefficient vector and optimal parameter settings for a dictionary with parameterized atoms or molecules.

A. Joint Dictionary and Sparse Coefficient Optimization

We begin with a dictionary whose atoms depend on a set of parameters $\boldsymbol{\eta}$; each parameter may or may not be shared by atoms or molecules. Furthermore, we consider the ℓ_p relaxation to the sparse signal representation problem mentioned in Section II-B [19]. The optimization problem at hand then is to minimize the following cost function:

$$J(\mathbf{a}, \boldsymbol{\eta}) = \|\mathbf{g} - \boldsymbol{\Phi}(\boldsymbol{\eta})\mathbf{a}\|_2^2 + \alpha\|\mathbf{a}\|_p^p, \quad p < 1, \qquad (7)$$

jointly determining a dictionary $\boldsymbol{\Phi}(\boldsymbol{\eta})$ and coefficients $\mathbf{a}$.

To carry out the joint minimization, we take a coordinate descent approach, alternately optimizing over the coefficients and dictionary parameters. The two optimizations are:

$$\mathbf{a}^{(t+1)} = \arg\min_{\mathbf{a}} \left\|\mathbf{g} - \boldsymbol{\Phi}\left(\boldsymbol{\eta}^{(t)}\right)\mathbf{a}\right\|_2^2 + \alpha\|\mathbf{a}\|_p^p. \qquad (8)$$

$$\boldsymbol{\eta}^{(t+1)} = \arg\min_{\boldsymbol{\eta}} \left\|\mathbf{g} - \boldsymbol{\Phi}(\boldsymbol{\eta})\mathbf{a}^{(t+1)}\right\|_2^2 + \alpha\left\|\mathbf{a}^{(t+1)}\right\|_p^p$$
$$= \arg\min_{\boldsymbol{\eta}} \left\|\mathbf{g} - \boldsymbol{\Phi}(\boldsymbol{\eta})\mathbf{a}^{(t+1)}\right\|_2^2. \qquad (9)$$

The application will guide the particular initialization for $\boldsymbol{\eta}$. The non-convex minimization (8) may be performed using the graph-structured algorithms of Section II and Section III, or using quasi-Newton optimization [19].

The minimization (9) may be recognized as nonlinear least-squares; many techniques exist in the literature including the trust-region reflective Newton algorithm that we use [36]. Linear inequality constraints on the parameter vector $\boldsymbol{\eta}$ may be handled within this framework. Termination of the procedure is upon the change in $\boldsymbol{\eta}$ falling below a small constant.

B. Characterization of Migratory Scattering Centers

We demonstrate joint dictionary parameter and sparse representation optimization on the characterization of a phenomenon in wide-angle SAR imaging different from anisotropy. Certain scattering mechanisms migrate as a function of aspect angle θ in wide-angle imaging [37], [38]. Migration occurs when radar signals bounce back from the closest surface of a physical object, but the closest surface of the object is different from different viewing angles; the physical object is not really moving, but appears to move in the measurement domain. By accounting for this effect in solving the inverse problem, a physically meaningful, parsimonious description can be extracted.

For example, considering a circular cylinder, the point of reflection on the surface closest to the radar can be parameterized as a function of θ around the center of the cylinder (x_c, y_c) using the radius of the cylinder η. When $\theta = 0$, the scatterer appears to be at $(x_c - \eta, y_c)$, which we define as $(\bar{x}, \bar{y})$. The observation model for migratory point scatterers is:

$$g(f, \theta) = \sum_{l=1}^{L} s_l(\theta) e^{-j\frac{4\pi f}{c}((\bar{x}_l + \eta_l)\cos\theta + \bar{y}_l \sin\theta - \eta_l)}. \quad (10)$$

A dictionary expansion for the observation model is:

$$g(f, \theta) = \sum_{l=1}^{L} \sum_{m=1}^{M} a_{lm} b_m(\theta) e^{-j\frac{4\pi f}{c}((\bar{x}_l + \eta_l)\cos\theta + \bar{y}_l \sin\theta - \eta_l)}. \quad (11)$$

In this instance, the atoms are parameterized by the radius η, and moreover, all atoms in molecule l share a common radius η_l; hence $\boldsymbol{\eta}$ is an L-vector of parameters. The inverse problem is to jointly recover the anisotropy and radius of migration of all scatterers in the ground patch.

The radius is constrained to be non-negative, i.e. $\boldsymbol{\eta} \geq \mathbf{0}$. Most scatterers are not migratory, and thus we initialize $\boldsymbol{\eta}$ with all zeroes. Often in practice, the coefficient vector $\mathbf{a}$ retains its sparsity structure on every iteration because even for $\boldsymbol{\eta} = \mathbf{0}$, characterized anisotropy may be close to correct, or at least have the correct support. The procedure may be envisioned as simultaneously inflating L balloons.

As an example, we look at data from XPatch of a scene containing a tophat that exhibits circular migratory scattering. In the aperture with $N = 15$ aspect angles spaced one degree apart, the tophat also has anisotropy, as shown in Fig. 11a. The magnitudes as well as the real and imaginary parts of the measurements are shown, as migratory scattering affects phase, not magnitude. An image of the scene formed using the polar format algorithm, a conventional method based on the inverse Fourier transform, is shown in Fig. 11b.

After identifying the spatial location with largest magnitude in the conventionally formed image, the coordinate descent described in this section is applied with $L = 1$. A raised triangle shape is used for the atoms. The solution has radius 5.314 meters and anisotropy as plotted in Fig. 11a. The circular migration of radius 5.314 meters is overlaid on and matches well with the conventional image in Fig. 11b. Coordinate descent to jointly optimize over radius and anisotropy is effective
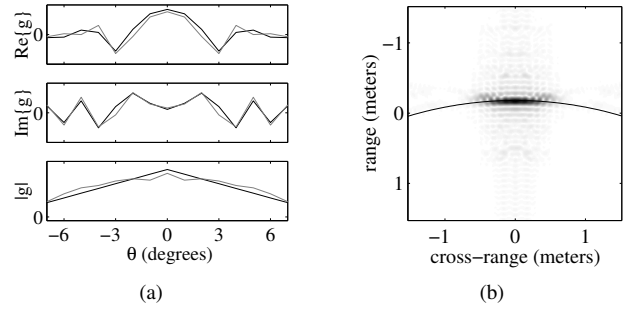

Fig. 11. Tophat example: (a) aspect-dependent scattering measurement (gray line) and solution (black line); (b) conventionally formed image with migration solution overlaid.

with realistic data seen here, and with several scatterers in a scene ($L > 1$), see [25]. By allowing for a non-zero radius, image formation is not simply pixel-based but more region-based. Although point scatterers can be equated to spatial locations, if information about migration is considered, the scatterer is more of an object-level construct.

We have looked at characterizing the migration of scatterers when the migration is circular in shape. Circles are an important subset of migratory scattering because many man-made objects contain scatterers with circular migration. However, any shape defined by a radius function $\eta(\theta)$ around a center is easily expressed in the observation model:

$$g(f, \theta) = \sum_{l=1}^{L} s_l(\theta) e^{-j\frac{4\pi f}{c}((\bar{x}_l + \eta_l(0))\cos\theta + \bar{y}_l \sin\theta - \eta_l(\theta))}. \quad (12)$$

Under this model, η_l is not constant across all angles, so a length L vector of parameters is not sufficient. One option is to take a functional form for $\eta_l(\theta)$ with more degrees of freedom than just a constant function, such as a polynomial, and lengthen the parameter vector $\boldsymbol{\eta}$. Another option is to locally, i.e. in small segments of θ, approximate $\eta_l(\theta)$ with pieces of circles [25]. The phenomenon of migratory scattering, which has rarely been explored in the literature, is a source of information that can be mined for details about object shape and size.

V. SIMULTANEOUS SPARSITY IN MULTIPLE DOMAINS

In the previous sections, we use an overcomplete dictionary $\boldsymbol{\Phi}$ to represent a signal $\mathbf{g}$, assuming that a sparse representation exists and then finding it. Our assumption in those sections is that $\mathbf{g}$ is sparse in the domain of the atoms. In this section, reverting to a known and fixed dictionary, we look at signals that are sparse in the domain of that known and fixed dictionary, but are also sparse in one or more other domains. The goal is to develop a formulation that recovers parsimonious representations, semantically interpretable in the case of inverse problems, making use of sparsity in all domains. Note that in the end, solutions will still be representations in terms of the atoms of the dictionary.
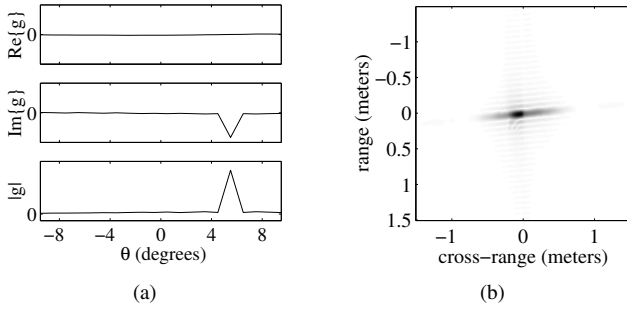
Fig. 12. Glint example: (a) aspect-dependent scattering measurement; (b) conventionally formed image.

A. Additional Sparsity Terms

For sparsity in the domain of the dictionary, the ℓ_p relaxation as an objective function is:

$$J(\mathbf{a}) = \|\mathbf{g} - \mathbf{\Phi}\mathbf{a}\|_2^2 + \alpha\|\mathbf{a}\|_p^p, \quad p < 1. \tag{13}$$

Let us assume that $\mathbf{g}$ is also sparse in a transformed domain and that that sparsity is to be exploited as well. First note that taking an orthonormal transformation of both the signal $\mathbf{g}$ and dictionary $\mathbf{\Phi}$ does not change the cost function. Also, the dictionary $\mathbf{\Phi}$ is fixed; consequently, we keep the data fidelity term as is, and append additional sparsity terms.

$$J(\mathbf{a}) = \|\mathbf{g} - \mathbf{\Phi}\mathbf{a}\|_2^2 + \sum_i \alpha_i \|\mathbf{F}_i(\mathbf{a})\|_p^p. \tag{14}$$

The functions $\mathbf{F}_i(\mathbf{a})$ return vectors related to the domain in which sparsity is to be favored. For the domain of the dictionary atoms, $\mathbf{F}_i$ is an identity operation. For domains that are transformations of the original domain, $\mathbf{F}_i$ is constructed as follows.

The operation $\mathbf{F}_i$ is the composition of three simpler operations. First, since the coefficients themselves have no particular meaning until paired with their corresponding atoms, initially $\mathbf{F}_i$ takes the coefficients through the atoms ϕ_m. Thereafter, the second operation is transformation to another domain. Finally, further operations in the transformed domain may follow. If all $\mathbf{F}_i(\mathbf{a})$ are linear, i.e. matrix-vector products, then the cost function may be optimized using quasi-Newton optimization [19] or the graph-structured algorithm using quasi-Newton optimization in each iteration. A concrete application given below constructs such $\mathbf{F}_i$.

B. Parsimonious Representation Recovery of Glint Anisotropy

Scattering behavior known as glint is produced by long, flat metal plates and is not migratory, has very narrow anisotropy, and corresponds to a line segment in the x-y domain oriented at the same angle as the center angle of the anisotropy. Fig. 12a shows aspect-dependent scattering of glint anisotropy from XPatch data and Fig. 12b shows a conventionally formed image. A parsimonious representation ought to explain scattering with a single scattering center, not with a collection of scatterers located on the line segment. We apply the formulation (14) both to favor sparsity among atoms and to favor sparsity along lines [38].

To favor sparsity among atoms, $\mathbf{F}_1$ is the identity. We now find a domain in which sparsity along a line can be favored. The normal parameter space of the Hough transform, the ρ-θ plane, and image space, the x-y plane, are related by the property that a set of points lying on the same line in image space corresponds to a set of sinusoids that intersect at a common point in parameter space [39]. Thus sparsity among scatterers in individual ρ-θ cells achieves the goal of sparsity among points on a line.

In [40], a Hough space sparsifying regularization approach is employed to enhance and detect straight lines in positive real-valued images by imposing sparsity when taking the image data to the ρ-θ plane. Parameter space cells with small counts are suppressed and cells with large counts are enhanced; thus, non-line features are suppressed and line features are enhanced in image space. The goals in our work are different and consequently, the sparsity terms are of a different flavor as well.

The range profile domain in SAR, a one-dimensional inverse Fourier transform of the phase history measurement domain, is equivalent to the parameter space of the Hough transform. It follows that for sparsity among scatterers in cell (ρ_k, θ_n), a sparsity term of the form $\||\mathbf{L}_{kn}\mathbf{a}|\|_p^p$ is used, where $\mathbf{L}_{kn}$ is a linear operator that is a composition of a block-diagonal version of the dictionary to bring the coefficients to the phase history domain, a discrete Fourier transform operator to go to the range profile domain, and a selection operator to select cell (ρ_k, θ_n). The resulting vector $\mathbf{L}_{kn}\mathbf{a}$ is of length L. Favoring sparsity in all range profile cells, the overall sparsity cost function is:

$$J(\mathbf{a}) = \|\mathbf{g} - \mathbf{\Phi}\mathbf{a}\|_2^2 + \alpha_1 \|\mathbf{a}\|_p^p + \alpha_2 \sum_{k=1}^K \sum_{n=1}^N \||\mathbf{L}_{kn}\mathbf{a}|\|_p^p. \tag{15}$$

The parameters α_1 and α_2 control the influence of the two sparsity terms. When $\alpha_2 = 0$, the cost function reduces to (13).

We solve the inverse problem with $L = 24$ pixels of interest identified by having large magnitude in the conventional image Fig. 12b. These 24 pixels are along a diagonal line more or less. The measurements are at $N = 20$ aspect angles over a $19°$ aperture with the glint at $5.5°$.

Let us define two counts related to the sparsity of the solution and look at their behavior as α_1 and α_2 are varied. We define L_A as the number of molecules out of the possible $L = 24$ that have at least one non-zero coefficient in the solution. Also, M_A is defined as the average number of non-zero coefficients per molecule in those molecules that have at least one non-zero coefficient. The maximum possible value of M_A is M, which is 210 for $N = 20$. When L_A is zero, M_A is defined to be zero. Solutions are obtained using the quasi-Newton method to minimize (15).

The two counts L_A and M_A are given in Table I for different values of α_1 and α_2. First, it should be noted that when α_1 and α_2 get too large, all of the coefficients go to zero. The main thing to take note of is that when $\alpha_2 = 0$, $L_A = 24$, i.e. all spatial locations provide contributions to the solution, but as α_2 increases, sparsity along a line is a greater influence and the number of contributing spatial locations decreases to one.

TABLE I

L_A AND M_A AS A FUNCTION OF THE PARAMETERS α_1 AND α_2.

$(L_\mathrm{A}, M_\mathrm{A})$	$\alpha_1 = 0$	$\alpha_1 = 10$	$\alpha_1 = 20$	$\alpha_1 = 30$	$\alpha_1 = 40$
$\alpha_2 = 0$	$(24, 39)$	$(24, 1)$	$(24, 1)$	$(24, 1)$	$(0, 0)$
$\alpha_2 = 5$	$(2, 39)$	$(3, 1)$	$(3, 1)$	$(3, 1)$	$(0, 0)$
$\alpha_2 = 10$	$(1, 39)$	$(2, 1)$	$(4, 1)$	$(2, 1)$	$(0, 0)$
$\alpha_2 = 15$	$(1, 39)$	$(4, 1)$	$(3, 1)$	$(1, 1)^*$	$(0, 0)$
$\alpha_2 = 20$	$(1, 39)$	$(4, 1)$	$(3, 1)$	$(1, 1)^*$	$(0, 0)$
$\alpha_2 = 25$	$(1, 39)$	$(4, 1)$	$(2, 1)$	$(0, 0)$	$(0, 0)$
$\alpha_2 = 30$	$(1, 39)$	$(2, 1)$	$(1, 1)^*$	$(0, 0)$	$(0, 0)$
$\alpha_2 = 35$	$(1, 39)$	$(1, 1)^*$	$(1, 1)^*$	$(0, 0)$	$(0, 0)$
$\alpha_2 = 40$	$(1, 39)$	$(1, 1)^*$	$(1, 1)^*$	$(0, 0)$	$(0, 0)$

Fig. 13. Coefficient magnitudes in 8-level guiding graph as signal **g** is varied from coarse to fine.

Sparsity among atoms is not enough for the solution on XPatch data to be parsimonious in the number of spatial locations, sparsity along a line is also required.

It can be seen that when $\alpha_1 = 0$, 39 atoms per spatial location contribute, not very sparse. For larger α_1, just one atom per spatial location contributes. Considering the behavior of L_A and M_A together, we note that the two sparsity terms are fairly orthogonal; the main effect of sparsity among atoms is on the number of atoms per spatial location and the main effect of sparsity along a line is on the number of spatial locations, as per the design objective.

A sparse and physically interpretable approximation ought to assign all of the scattering to the leaf atom at $5.5°$ of a single spatial location. Such a solution with one non-zero coefficient is recovered for the (α_1, α_2) pairs marked with an asterisk in Table I.

Through the example it has been seen that both types of sparsity are necessary to recover a solution that represents the scattering as coming from a single point and with very thin anisotropy explained by a single atom. With this representation, spatial properties about the object being imaged, such as orientation and physical extent, may be inferred. Although the same object-level inferences could have been made with $\alpha_2 = 0$, in that case, L such objects would be indicated rather than one, which does not make physical sense. Points have more meaning than just pixels with aspect-dependent amplitudes.

VI. CONCLUSION

We looked at methods of obtaining sparse signal representations and approximations from overcomplete dictionaries with hierarchical structures within subdictionaries, focusing on the context of coherent inverse problems with physically interpretable dictionary elements. We developed a heuristic method of solution for such problems that takes advantage of the structure by relating the problem to search on graphs. We also took a step back from the classic sparse signal representation problem to consider dictionary refinement as well as obtaining solutions simultaneously sparse in multiple domains. Under dictionary refinement, a coordinate descent approach was developed to jointly optimize parameterized atoms and coefficients, whereas under simultaneous sparsity, an extended sparsifying cost function was minimized.

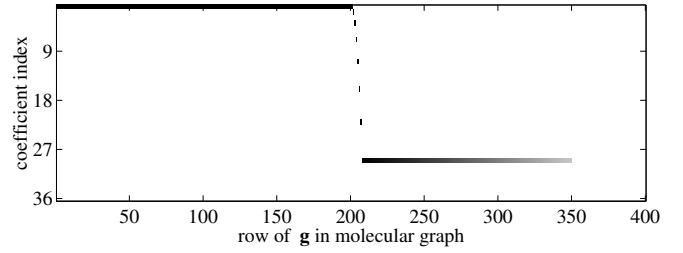The methods were demonstrated on various facets of wide-angle SAR, but are general enough to transfer to other applications with appropriate dictionaries. In the SAR context, starting from the same low-level measurements used by conventional image formation techniques, we have taken a step farther in scene understanding while also taking into account phenomena such as anisotropy that cause inaccuracies in conventional methods. We have started to move away from a pixel representation to more of an object-level representation through the use of a physically meaningful dictionary.

APPENDIX

Two experimental results are given as empirical validation for the search heuristic and stopping criterion described in Section II-B. We show that solutions from subdictionaries do in fact have non-zero coefficients for atoms most 'similar' to the signal **g**, particularly when **g** is not contained in the subdictionary. For the experiments, the molecular graph has $N = 400$ levels and the guiding graph has $G = 8$ levels. Keeping the guiding graph fixed within the molecular graph, the behavior of the solution **a** is observed as the signal **g** is varied. Quasi-Newton optimization is used to obtain the sparse solution coefficients **a**.

In the first experiment, with results in Fig. 13, the guiding graph is fixed with root at the left-most node of level 200 in the molecular graph. The true signal **g** is varied from coarse to fine support. In terms of the molecular graph, the true coefficient is varied, starting at the root node, through all nodes along the left edge of the graph, to the left-most node of level 400. In the plot, the row in the molecular graph which contains **g** is plotted on the horizontal axis. The magnitudes of the 36 coefficients in **a** are indicated by shading (white is zero); each horizontal strip is for one of the coefficients. Most coefficients are zero for all **g** due to sparsity. In the regime where the guiding graph is below the true coefficient, the coefficient of the guiding graph root node is non-zero. In the regime where the guiding graph covers the true coefficient, the correct coefficient is non-zero. When the guiding graph is above the true coefficient, the coefficient of the bottom left node, the node in the last level closest to the truth, is non-zero and others are zero. It should be noted that the influence of the finest signals does not reach up to make any guiding graph node coefficients non-zero (a consequence of regularization).

In the experiment yielding the results of Fig. 14, the guiding graph is fixed with root at the center node of level 200 instead of the left-most node. The true node is varied from left to right across the molecular graph at level 210, three levels below
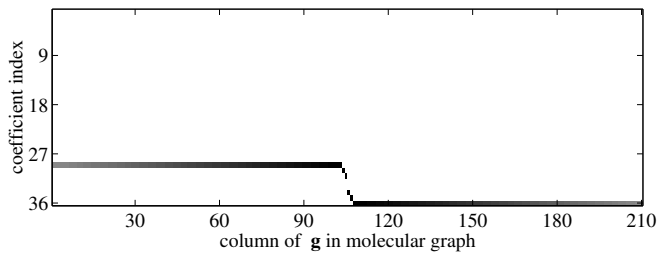
Fig. 14. Coefficient magnitudes in 8-level guiding graph as signal **g** is shifted from left to right.

the bottom of the guiding graph. This figure is organized in the same manner as Fig. 13, but the horizontal axis indicates the column of **g** in the molecular graph. From these results, first it is apparent that only coefficients in the last level of the guiding graph are non-zero, reconfirming results from the previous experiment. Second, it can be seen that when the truth is to the left of the guiding graph, the left-most node of level G is non-zero. Similarly, when the truth is to the right, the right node is non-zero; when the truth is underneath the 8-level graph, nodes in the interior of the last level are non-zero.

ACKNOWLEDGMENT

REFERENCES

[1] D. M. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[2] J.-J. Fuchs and B. Delyon, "Minimum L_1-norm reconstruction function for oversampled signals: Applications to time-delay estimation," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1666–1673, July 2000.

[3] M. Çetin, W. C. Karl, and A. S. Willsky, "Feature-preserving regularization method for complex-valued inverse problems with application to coherent imaging," *Optical Engineering*, vol. 45, no. 1, p. 017003, Jan. 2006.

[4] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Processing*, vol. 48, no. 3, pp. 600–616, Mar. 1997.

[5] B. D. Jeffs and M. Gunsay, "Restoration of blurred star field images by maximally sparse optimization," *IEEE Trans. Image Processing*, vol. 2, no. 2, pp. 202–211, Apr. 1993.

[6] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Computing*, vol. 20, no. 1, pp. 33–61, Aug. 1998.

[8] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. National Acad. Sciences*, vol. 100, no. 5, pp. 2197–2202, Mar. 4 2003.

[9] D. M. Malioutov, M. Çetin, and A. S. Willsky, "Optimal sparse representations in general overcomplete bases," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 2, Montréal, May 2004, pp. 793–796.

[10] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[11] L. Daudet, "Sparse and structured decompositions of signals with the molecular matching pursuit," *IEEE Trans. Audio Speech Language Processing*, vol. 14, no. 5, pp. 1808–1816, Sept. 2006.

[12] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Signal Processing*, vol. 51, no. 1, pp. 101–111, Jan. 2003.

[13] L. Granai and P. Vandergheynst, "Sparse decomposition over multi-component redundant dictionaries," in *Proc. IEEE Multimedia Signal Processing Workshop*, Siena, Italy, Sept. 2004, pp. 494–497.

[14] A. Shoa and S. Shirani, "Tree structure search for matching pursuit," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, Genoa, Italy, Sept. 2005, pp. 908–911.

[15] C. La and M. N. Do, "Signal reconstruction using sparse tree representations," *Proc. SPIE*, vol. 5914, p. 591410W, Sept. 2005.

[16] P. Jost, P. Vandergheynst, and P. Frossard, "Tree-based pursuit: Algorithm and properties," *IEEE Trans. Signal Processing*, vol. 54, no. 12, pp. 4685–4697, Dec. 2006.

[17] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser, "Signal processing and compression with wave packets," in *Proc. Int. Conf. Wavelets*, Marseille, France, May 1989.

[18] S. Jaggi, W. C. Karl, S. Mallat, and A. S. Willsky, "High resolution pursuit for feature extraction," *J. Appl. Comp. Harmonic Analysis*, vol. 5, pp. 428–449, 1998.

[19] M. Çetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Trans. Image Processing*, vol. 10, no. 4, pp. 623–631, Apr. 2001.

[20] D. C. Munson, Jr., J. D. O'Brien, and W. K. Jenkins, "A tomographic formulation of spotlight-mode synthetic aperture radar," *Proc. IEEE*, vol. 71, no. 8, pp. 917–925, Aug. 1983.

[21] R. L. Moses, L. C. Potter, and M. Çetin, "Wide angle SAR imaging," *Proc. SPIE*, vol. 5427, pp. 164–175, Apr. 2004.

[22] J. B. Keller, "Geometrical theory of diffraction," *J. Opt. Soc. Amer.*, vol. 52, no. 2, pp. 116–130, Feb. 1962.

[23] F. G. Meyer, A. Z. Averbuch, and R. R. Coifman, "Multilayered image representation: Application to image compression," *IEEE Trans. Image Processing*, vol. 11, no. 9, pp. 1072–1080, Sept. 2002.

[24] K. R. Varshney, M. Çetin, J. W. Fisher, III, and A. S. Willsky, "Joint image formation and anisotropy characterization in wide-angle SAR," *Proc. SPIE*, vol. 6237, p. 62370D, Apr. 2006.

[25] K. R. Varshney, "Joint anisotropy characterization and image formation in wide-angle synthetic aperture radar," Master's thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2006.

[26] "Backhoe data dome and Visual-D challenge problem," Available at Air Force Research Laboratory Sensor Data Management System (https://www.sdms.afrl.af.mil/main.php), 2004.

[27] M. A. Koets and R. L. Moses, "Feature extraction using attributed scattering center models on SAR imagery," *Proc. SPIE*, vol. 3721, pp. 104–115, Apr. 1999.

[28] L. C. Potter and R. L. Moses, "Attributed scattering centers for SAR ATR," *IEEE Trans. Image Processing*, vol. 6, no. 1, pp. 79–91, Jan. 1997.

[29] L. C. Trintinalia, R. Bhalla, and H. Ling, "Scattering center parameterization of wide-angle backscattered data using adaptive Gaussian representation," *IEEE Trans. Antennas Propagat.*, vol. 45, no. 11, pp. 1664–1668, Nov. 1997.

[30] R. D. Chaney, A. S. Willsky, and L. M. Novak, "Coherent aspect-dependent SAR image formation," *Proc. SPIE*, vol. 2230, pp. 256–274, Apr. 1994.

[31] L. R. Flake, S. C. Ahalt, and A. K. Krishnamurthy, "Detecting anisotropic scattering with hidden Markov models," *IEE P. Radar, Son., Nav.*, vol. 144, no. 2, pp. 81–86, Apr. 1997.

[32] A. J. Kim, J. W. Fisher, III, and A. S. Willsky, "Detection and analysis of anisotropic scattering in SAR data," *Multidimensional Systems and Signal Processing*, vol. 14, no. 1-3, pp. 49–82, Jan. 2003.

[33] M. Çetin and R. L. Moses, "SAR imaging from partial-aperture data with frequency-band omissions," *Proc. SPIE*, vol. 5808, pp. 32–43, Mar. 2005.

[34] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T.-W. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Comp.*, vol. 15, no. 2, pp. 349–396, Feb. 2003.

[35] M. Aharon, M. Elad, and A. M. Bruckstein, "On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them," *Linear Algebra Appl.*, vol. 416, no. 1, pp. 48–67, July 2006.

[36] T. F. Coleman and Y. Li, "A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables," *SIAM J. Optimization*, vol. 6, no. 4, pp. 1040–1058, 1996.

[37] R. Bhalla, A. M. Raynal, H. Ling, J. Moore, and V. J. Velten, "Angular description for 3D scattering centers," *Proc. SPIE*, vol. 6237, p. 623706, Apr. 2006.

[38] K. R. Varshney, M. Çetin, J. W. Fisher, III, and A. S. Willsky, "Wide-angle SAR image formation with migratory scattering centers and regularization in Hough space," in *Proc. Adaptive Sensor Array Processing Workshop*, Lexington, Massachusetts, June 2006.

[39] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Comm. ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972.

[40] N. Aggarwal and W. C. Karl, "Line detection in images through regularized Hough transform," *IEEE Trans. Image Processing*, vol. 15, no. 3, pp. 582–591, Mar. 2006.

Hyper-parameter selection in non-quadratic regularization-based radar image formation

Özge Batu[a] and Müjdat Çetin[a,b]

[a]Faculty of Engineering and Natural Sciences,
Sabancı University, Orhanlı, Tuzla 34956 İstanbul, Turkey

[b]Laboratory for Information and Decision Systems,
Massachusetts Institute of Technology, Cambridge, MA 02139, USA

ABSTRACT

We consider the problem of automatic parameter selection in regularization-based radar image formation techniques. It has previously been shown that non-quadratic regularization produces feature-enhanced radar images; can yield superresolution; is robust to uncertain or limited data; and can generate enhanced images in non-conventional data collection scenarios such as sparse aperture imaging. However, this regularized imaging framework involves some hyper-parameters, whose choice is crucial because that directly affects the characteristics of the reconstruction. Hence there is interest in developing methods for automatic parameter choice. We investigate Stein's unbiased risk estimator (SURE) and generalized cross-validation (GCV) for automatic selection of hyper-parameters in regularized radar imaging. We present experimental results based on the Air Force Research Laboratory (AFRL) "Backhoe Data Dome," to demonstrate and discuss the effectiveness of these methods.

Keywords: synthetic aperture radar, hyper-parameter selection, sparse-aperture imaging, feature-enchanced imaging, inverse problems

1. INTRODUCTION

Conventional image formation techniques for synthetic aperture radar (SAR) suffer from low resolution, speckle and sidelobe artifacts. These effects pose challenges for SAR images when used in automatic target detection and recognition tasks. Recently, new SAR image formation algorithms have been proposed to produce high quality images which provide increased resolution and reduced artifacts [1,2,3]. We consider the non-quadratic regularization-based approach of [1] which aims at providing feature-enhanced SAR images. The idea behind this approach is to emphasize appropriate features by means of regularizing the solution. In fact, regularization methods are well known and widely used for real-valued image restoration and reconstruction problems. However SAR imaging involves some difficulties in application of these methods. As an example, SAR involves complex-valued reflectivities. Considering and addressing such difficulties, extensions of real-valued non-quadratic regularization methods have been developed for SAR imaging.

Regularization methods, in general, try to balance the fidelity to data and prior knowledge to obtain a stable solution. This stability is ensured through a scalar parameter which is called regularization parameter or hyper-parameter. Selection of this parameter is another problem in a regularization framework. There exist several approaches which have been mostly practised in quadratic regularization methods such as Tikhonov regularization. Recently, non-quadratic methods have acquired greater importance thanks to their property of preserving useful features such as edges. Hence there is interest in developing methods for automatic parameter choice in the non-quadratic setting.

_____

We consider Stein's unbiased risk estimator (SURE) and generalized cross-validation (GCV) for parameter selection in non-quadratic regularization-based radar imaging. They have been both used in problems with quadratic constraints [4,5] but the experiments for non-quadratic methods are limited [6]. We propose their use in the regularization-based SAR image formation framework. We present the effectiveness of SURE and GCV through our experiments based on the Air Force Research Laboratory (AFRL) "Backhoe Data Dome" [7].

2. REGULARIZATION-BASED SAR IMAGING

For feature-enhanced image formation we consider an approach based on non-quadratic regularization [1]. The framework of [1] relies on the SAR observation process expressed in the following form:

$$y = Hf + w \tag{1}$$

where H represents a complex-valued discrete SAR operator, w stands for additive noise, y and f are data and the reflectivity field, respectively. Here we prefer to use the conventional image as the input data, hence the technique works as a deconvolution method. In this framework, SAR image reconstruction problem is formulated as the following optimization problem

$$\hat{f} = \arg\min_f J(f). \tag{2}$$

One choice for $J(f)$, which we consider here, has the following form:

$$J(f) = \|y - Hf\|_2^2 + \lambda \|f\|_p^p \tag{3}$$

where $\|.\|_p^p$ denotes the ℓ_p-norm and λ is a scalar parameter. The first term in the objective function (3) is a data fidelity term, which incorporates the SAR observation model (1), and thus information about the observation geometry. The second term in (3) incorporates prior information reflecting the nature of the field f, and is aimed at enhancing point-based features. Additional terms like a smoothness penalty on f can be employed in this framework to emphasize other characteristics of the field. However, in fact, many object recognition methods rely on locations of dominant point scatterers extracted from SAR images. Therefore we choose the cost function $J(f)$ to be as in (3) throughout this work, and thus produce images in which point-based features are enhanced. It has been known that minimum ℓ_p-norm reconstruction with $p \leq 1$ provides localized energy concentrations in the resultant image. In such images, most elements are forced to be small, on the other hand, a few are allowed to have very large values. With respect to ℓ_2-norm reconstruction this approach favors a field with smaller number of dominant scatterers. This type of constraint aims to suppress artifacts and increase the resolvability of scatterers.

To avoid the problems due to nondifferentiability of the objective function around the origin, a smooth approximation to the ℓ_p-norm is used, and the objective function takes the following form

$$J(f) = \|y - Hf\|_2^2 + \lambda \sum_{i=1}^{n} \left(|f_i|^2 + \beta \right)^{p/2} \tag{4}$$

where f_i denotes the ith element of f, n is the number of pixels in f, and β is a small scalar. The estimate $\hat{f}$ is the solution of the following equation:

$$\hat{f} = \left(H^T H + \lambda W_\beta(\hat{f}) \right)^{-1} H^T y \tag{5}$$

where $W_\beta\left(\hat{f}\right)$ is a diagonal weight matrix whose ith diagonal element is $p\left(\left|f_i\right|^2 + \beta\right)^{(p-2)/2}$. The weight matrix acts in the following manner: if there is a scattering object in the field of interest then f_i's will be large, and thus corresponding elements of $W_\beta\left(\hat{f}\right)$ will be small and allowing large intensities. Otherwise, elements of $W_\beta\left(\hat{f}\right)$ will be large and suppress the energy concentrations at that location.

The expression in (5) is still nonlinear and no closed form solution exists. An iterative procedure can handle the nonlinearity and results in the formulation:

$$\hat{f}^{k+1} = \left(H^T H + \lambda W_\beta\left(\hat{f}^k\right)\right)^{-1} H^T y \tag{6}$$

where $\hat{f}^k$ is the estimate calculated in the kth iteration. In this way, the problem becomes linear at each individual step.

3. HYPER-PARAMETER SELECTION METHODS

The objective function in (4) contains a scalar parameter λ which has a role in determining the behavior of the reconstructed field $\hat{f}$. Small parameter values makes data fidelity term; i.e. first term in (4), dominate the solution whereas large values of λ impose greater importance to prior term and ensure that point-based features are enhanced. To choose λ in a data-driven way, we consider two methods: Stein's unbiased risk estimator (SURE) and generalized cross-validation (GCV).

3.1 Stein's unbiased risk estimator

Stein's unbiased risk estimator (SURE) is developed by Stein [4] for parameter selection in linear regression and it has been adapted for the solution of inverse problems. SURE aims to minimize the predictive risk:

$$\frac{1}{n}\left\|p_\lambda\right\|_2^2 = \frac{1}{n}\left\|H\hat{f}_\lambda - Hf\right\|_2^2 \tag{7}$$

which is basically mean squared norm of the predictive error. Here $\hat{f}_\lambda$ represents the solution obtained with parameter λ and f is the true, unknown reflectivity field. In fact, the predictive error is not computable since f is unknown, but it can be estimated using available information.

It has been shown that [4] an unbiased estimator of (7) is

$$U\left(\lambda\right) = \frac{1}{n}\left\|r\left(\lambda\right)\right\|_2^2 - \frac{2\sigma^2}{n}\sum_{i=1}^{n}\partial r_i\left(\lambda\right)/\partial y_i + \sigma^2 \tag{8}$$

where $r\left(\lambda\right) = H\hat{f}_\lambda - y$ and σ^2 is the varience of the noise w. $U\left(\lambda\right)$ is called Stein's unbiased risk estimator (SURE). It has been shown that [6] SURE takes the following form after some intermediate operations:

$$U\left(\lambda\right) = \frac{1}{n}\left\|r\left(\lambda\right)\right\|_2^2 + \frac{2\sigma^2}{n}trace\left(A_\lambda\right) - \sigma^2 \tag{9}$$

where $A_\lambda = HJ_{\hat{f}\hat{f}}^{-1}H^T$ and $J_{\hat{f}\hat{f}} = \partial^2 J/\partial\hat{f}^2$. For the cost function given in (4),

$$J_{\hat{f}\hat{f}} = H^T H + \lambda diag\left(p\left(\left|\hat{f}_i\right|^2 + \beta\right)^{(p-4)/2}\left(\left(p-1\right)\left|\hat{f}_i\right|^2 + \beta\right)\right). \tag{10}$$

The method chooses the parameter which minimizes $U(\lambda)$. Numerical results suggest that λ which minimizes the predictive risk will yield a small value for the estimation error of f. It is also noteworthy that SURE requires prior knowledge on the noise, namely its variance, in the model (1).

3.2 Generalized cross-validation

Another parameter selection method is generalized cross-validation (GCV) which is also an estimator for the predictive risk (7) and has the advantage of being independent of the prior knowledge on the noise w. The idea of GCV is as follows: choose λ such that the solution obtained in the presence of a missing data point predicts the missing point in a proper manner, when averaged over all ways of removing a point. The method intends to minimize the GCV function:

$$GCV(\lambda) = \frac{\frac{1}{n}\left\|r(\lambda)\right\|_2^2}{\left[\frac{1}{n}trace(I - A_\lambda)\right]^2} \tag{11}$$

where $r(\lambda)$ and A_λ are the same quantities in the SURE setting.

4. NUMERICAL OPTIMIZATION TOOLS

Both SURE and GCV involve computational difficulties when considered in the framework of non-quadratic regularization.[1] First of all, they require the computation of the matrix A_λ through large scale matrix multiplications and inversions which are not practical at all. Then they require the solution of an optimization problem over λ. To clarify the implementation details we discuss some numerical tools we use in the solution.

We note that all the matrix vector products in (6) are actually carried out by convolution operations (in the Fourier domain) such that there is no need to construct the convolution matrix H and deal with memory-intensive matrix operations. However convolutional operations do not help for evaluation of the GCV cost in (11) since it involves the trace of the matrix A_λ. Instead of calculating A_λ one can approximate the trace of A_λ by means of randomized trace estimation [8]. If q is a white noise vector with zero mean and unit variance, then an unbiased estimator for $trace(A_\lambda)$ can be constructed based on the random variable $t(\lambda) = q^T A_\lambda q$. The trace estimate can be computed as follows: first generate a number of independent realizations of q and compute $t(\lambda)$ for each, and then take the arithmetic mean of the random variable $t(\lambda)$ to be the trace estimate. In our experiments we also observed that the randomized trace estimate approaches successfully to the actual trace. Note that we compute $t(\lambda)$ using convolution operations without explicitly constructing the matrix A_λ. This method makes the computation of the GCV function feasible for a given λ. However there is still the issue of finding the minimizers of the GCV cost.

One way to find the minimum of the GCV cost is brute-force searching. After determining a reasonable range for values of λ, the range is divided into grids and solution is obtained for each grid point. Then, λ which gives the smallest function value is selected as the minimizer. Since this may yield extensive computations, appropriate optimization methods can be employed instead. Most of the methods are based on gradient information of the functions. However evaluation of the gradient of GCV appears to be a problem. Due to the complicated dependence of (9) and (11) on λ through $\hat{f}_\lambda$, it is not straightforward to compute the gradient. This difficulty leads us consider two different approaches: derivative-free optimization techniques and numerical computations of the gradient.

---

[1]In this section we only mention GCV for convenience, all the procedure is identical for SURE.

Golden section search is a one dimesional optimization method and does not require the gradient information [9]. This approach assumes that the function is unimodal (at least in an interval $[a,b]$), which is the case for GCV cost according to our observations. We first determine an interval $[a,b]$ for λ and choose two initial test points λ_1 and λ_2 in interval such that $\lambda_1 < \lambda_2$. Using those particular choices of λ's we find the solution $\hat{f}_\lambda$ and the GCV value for each. After evaluating GCV(λ_1) and GCV(λ_2) one of the following cases produces a new interval which is the subset of $[a,b]$.

Case 1: If GCV(λ_1) > GCV(λ_2), then the new interval is $[\lambda_1, b]$.

Case 2: If GCV(λ_1) ≤ GCV(λ_2), then the new interval is $[a, \lambda_2]$.

As a result, the interval is shrinked and two new test points are selected such that they divide the interval into the Golden section. Golden ratio requires:

$$\frac{\text{length of whole interval}}{\text{length of larger part of interval}} = \frac{\text{length of larger part of interval}}{\text{length of smaller part of interval}}$$

Thus the algorithm ends up with an interval of uncertainity; i.e. it does not provide a single point as the minimizer.

Since the challenge is the exact derivative computation we also consider approximating derivatives through finite differences. With an initial λ at hand we move to a close point $\lambda + \varepsilon$ and find the solution $\hat{f}_\lambda$ and the GCV value for each and calculate the finite difference of GCV at λ. Once the approximated gradient is found one can choose one of the line search algorithms to obtain the optimal point [10]. Since we don't want to deal with second order derivatives we choose a line search algorithm which uses a descent direction and thus does not require second order derivatives. To determine the step length one can use a step-length selection algorithm such as interpolation or use backtracking. In our experiments we apply the backtracking algorithm. The algorithm takes a step and checks the Armijo condition; if it is not satisfied, the step length is reduced and Armijo condition is checked again. The backtracking continues until a step length satisfying Armijo condition is found. Initial λ to be used in this procedure may be selected using a computationally less expensive method. For this task we employ the parameter selection approach in [11]. This method suggests the regularization parameter to be selected as $\lambda = \sigma\sqrt{2\log n}$ in a basis pursuit framework. Here σ is the standard deviation of the noise in (1) and n is the length of the noise vector.

5. EXPERIMENTS

We present 2D image reconstruction experiments based on the AFRL "Backhoe Data Dome and Visual-D challange problem" which consists of simulated wideband (7-13 GHz), full polarization, complex backscatter data from a backhoe vehicle in free space [7]. The backscatter data are available over a full upper 2π steradian viewing hemisphere. In our experiments, we use VV polarization data, centered at 10 GHz, and with an azimuthal span of 110° (centered at 45°). Advanced imaging strategies have enabled resolution-enhanced wide angle SAR imaging. We consider the point-enhanced composite imaging technique [12] and show experimental results in this framework. For composite imaging, we use 19 subapertures, with azimuth centers at 0°, 5°, …, 90°, and each with an azimuthal width of 20°. We consider two different bandwidths: 500 MHz and 1 GHz. For each of these bandwidths, we consider data with three different signal-to-noise ratios: 25 dB, 20 dB and 10 dB.

To be able to carry out some quantitative analysis, we have also created a synthetic problem which simulates imaging of a point-like scattering field in a narrow-angle imaging scenario. The field consists of five scatterers. We simulate SAR data with 1 GHz bandwidth and 25 dB SNR. We choose $p = 1$ in (3). The underlying true scene, the conventional reconstruction and reconstructed images with different regularization parameters are shown in Figure 1. In these images it is obviously seen that small parameter values are insufficient to enhance point-based features whereas large parameter values overregularize the solution and cause some scatterers to be unobservable. The image in Figure 1(d) is obtained
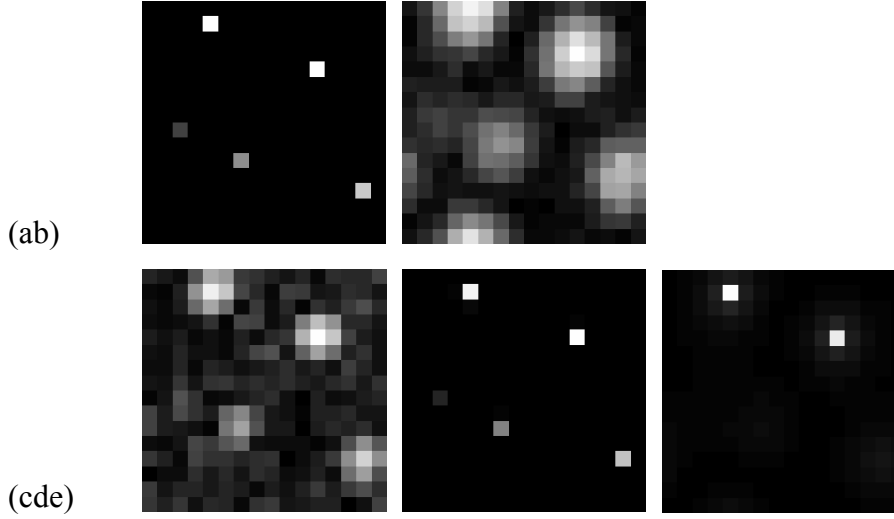
(ab)

(cde)

Figure 1. SAR images of a synthetic problem. (a) Underlying field consisting of point-like scatterers. (b) Conventional SAR image of the field. (c) Point-enhanced image with small λ. (d) Point-enhanced image with λ selected by GCV. (e) Point-enhanced image with large λ.

using λ selected by the GCV method. It appears to be an accurate reconstruction in the sense that it preserves all the five scatterers and does not cause significant artifacts.

In Figure 2, we show the SURE (solid, blue) and GCV (dashed, green) curves for the synthetic problem. The red (dash-dotted) curve indicates the estimation error which we define as $\frac{1}{n}\left\|\hat{f}-f\right\|_2^2$. The minimum point for each curve is enclosed by a square. Both SURE and GCV appear to be leading to slight under-regularization. Generally, SURE tends to choose a smaller λ value than GCV does. In fact, both functions are quite flat around the minimum; hence they are not very sensitive therein. Fortunately, the estimation error is not very sharp around the minimum either; and therefore SURE and GCV give reasonable results in terms of the estimation error. We also show the choice of the method proposed by Chen [11], and it behaves like an over-regularizer for this problem setting. It can be used as the initial parameter in the optimization procedure. However, we cannot be sure about the behaivour of this method for different settings since it depends only on the standard deviation of the noise and the size of the problem. For example; it will not respond to changes in the experimental scenario such as bandwidth (and in turn range resolution).

We now demonstrate the behavior of Golden section search and numerical gradient descent method for optimization. In Figure 3, we show the paths displaying the progress of the methods. For Golden section search algorithm we choose the initial interval the same as the interval we choose in brute-force searching ($\left[10^{-2},10^{0}\right]$). Golden section search progresses quite fast and ends up with an interval requiring small number of reconstructions. Finding an interval of uncertainity does not appear to be a trouble since SURE and GCV curves are quite flat around the minimum. In Figure 3(b) we show the progress of the line search algorithm based on numerical gradient computation. Red cross markers indicate the λ value for progressing iterations (For better visualization we do not show the points at each iteration; instead one of every three iteration points is marked). For both methods the most important thing is the number of evaluations of GCV and SURE. Let's consider a particular example, in which one would have to do 20 reconstructions to be able to determine λ with $\pm5\times10^{-2}$ variation in a brute-force search. We observe that it is possible to obtain the same precision with about 4 reconstructions in Golden search. Similarly, the numerical gradient computation-based algorithm provides almost the same advantage.
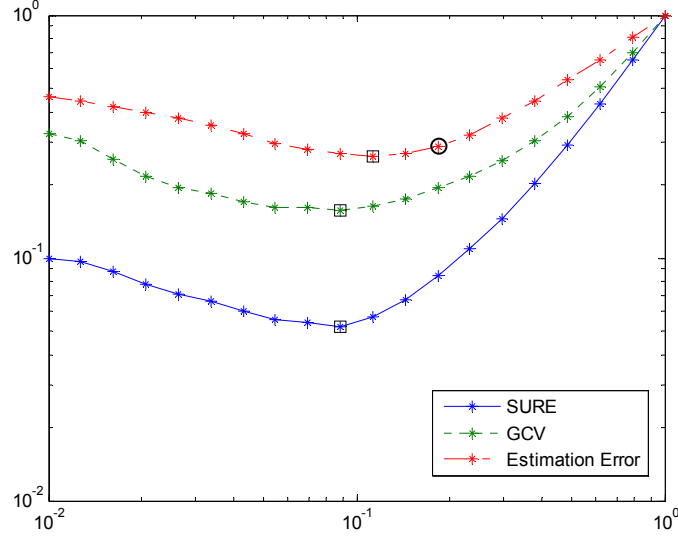
Figure 2. SURE, GCV and Estimation Error curves for the synthetic image consisting of point-like scatterers. Minimum points are enclosed by squares. The point enclosed by a circle is the parameter selected by the method proposed in [11].
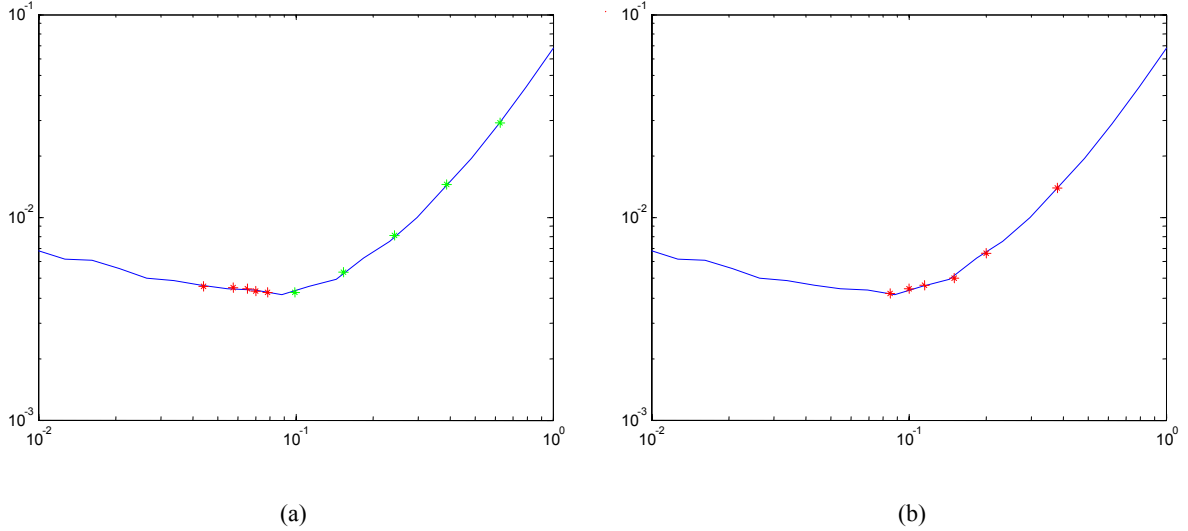

(a)　　　　　　　　　　　　　　　　　(b)

Figure 3. Paths for optimization methods. (a) Right (green) and the left (red) endpoints of the intervals for Golden section search. (b) Improvement steps in line search for numerically computed gradient descent..

For backhoe data we cannot demostrate the mean square error of estimated $\hat{f}_\lambda$ since the underlying field is not available a priori. Therefore we investigate the performance of the methods visually. First, we display the structure of SURE and GCV for backhoe data with bandwidth of 1 GHz and 500 MHz in Figure 4. A similar behavior to the synthetic problem is observed with this data. The curves are flat near the minimum, moreover the SURE curve is very flat for a wide range.

In Figure 5, we show the point-enhanced composite images obtained with different regularization parameters. Figure 5(c) is the image reconstructed using the parameter which is selected by GCV. Ideally, we would like to be able to observe the scattering centers of the backhoe in a good reconstruction. From this point of view GCV seems to serve the purpose. The under-regularized image in Figure 4(a) is dominated by artifacts and the over-regularized image in Figure 4(e) does not display the the structure of the backhoe correctly because of the unobservable scattering parts.
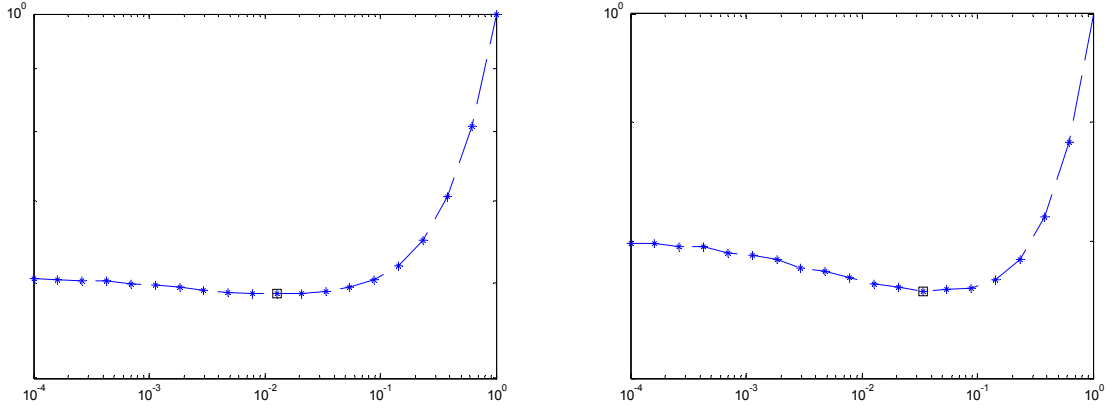
Figure 4. SURE (left) and GCV (right) curves for a subaperture image reconstructed from the data with SNR=20 dB and bandwidth of 1 GHz.
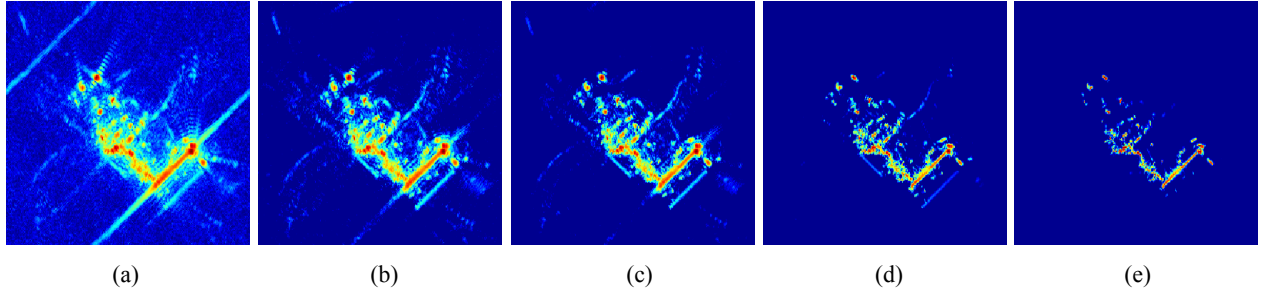

Figure 5. Feature-enhanced composite images using different λ's, bandwidth is 1 GHz and SNR=20 dB. λ_{GCV} denotes the parameter selected by GCV. (a) $\lambda=10^{-2}\lambda_{GCV}$. (b) $\lambda = 10^{-1}\lambda_{GCV}$. (c) $\lambda = \lambda_{GCV}$. (d) $\lambda = 10\lambda_{GCV}$. (e) $\lambda = 10^{2}\lambda_{GCV}$.

Different ℓ_p-norm's can be used in (3). A smaller value of p implies less penalty on large pixel values as compared to a larger p. This property favors a field with smaller number of dominant scatterers. This behavior can be observed in images obtained for different p values displayed in Figure 6.

In Figures 7, 8 and 9 we demonstrate the feature-enhanced composite images in the presence of noise. In these experiments we choose $p = 1$ in (3). We consider two bandwidths: 500 MHz and 1 GHz. It is possible to choose an
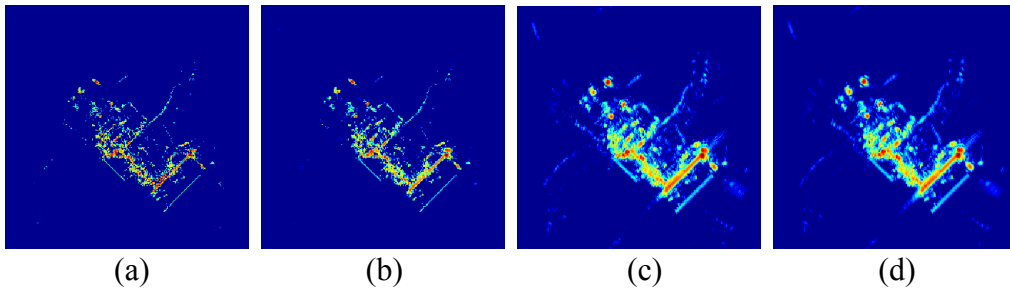

Figure 6. Composite, point-enhanced SAR images for p=0.6, 0.8, 1, 1.2. Regularization parameter λ is selected with GCV.
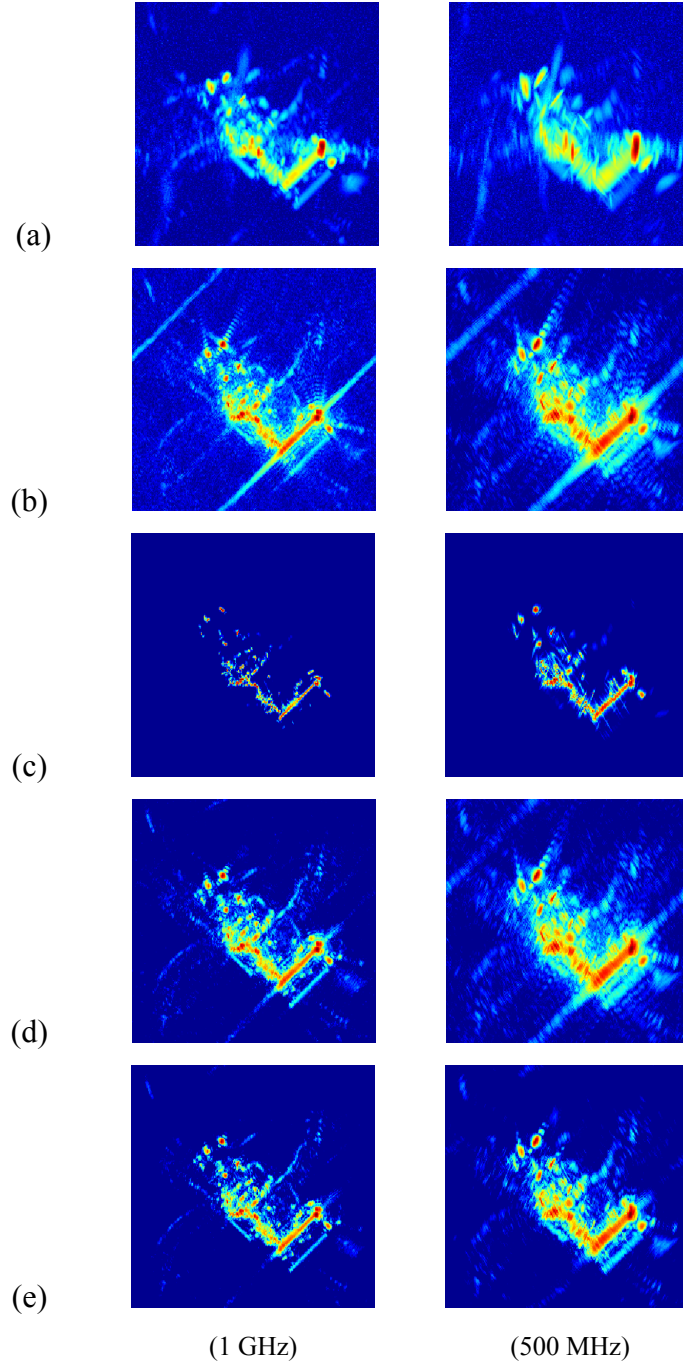
Figure 7. SAR images of the backhoe using bandwidths of 1 GHz and 500 MHz in the presence of SNR=25 dB. (a) Composite imaging. (b) Composite, point-enhanced imaging using small . (c) Composite, point-enhanced imaging using large λ. (d) Composite, point-enhanced imaging using λ selected by SURE. (e) Composite, point-enhanced imaging using λ selected by GCV.

individual λ for each subaperture seperately, however that would be computationaly very expensive. We have observed that the optimum λ does not vary significantly among different subapertures; thus we choose λ for one subaperture and use the same value for others. The results demonstrated in Figure 6 are obtained from the data with SNR=25 dB. The images in (a) are the conventional composite images. Images in (b) and (c) are obtained by very small and very large
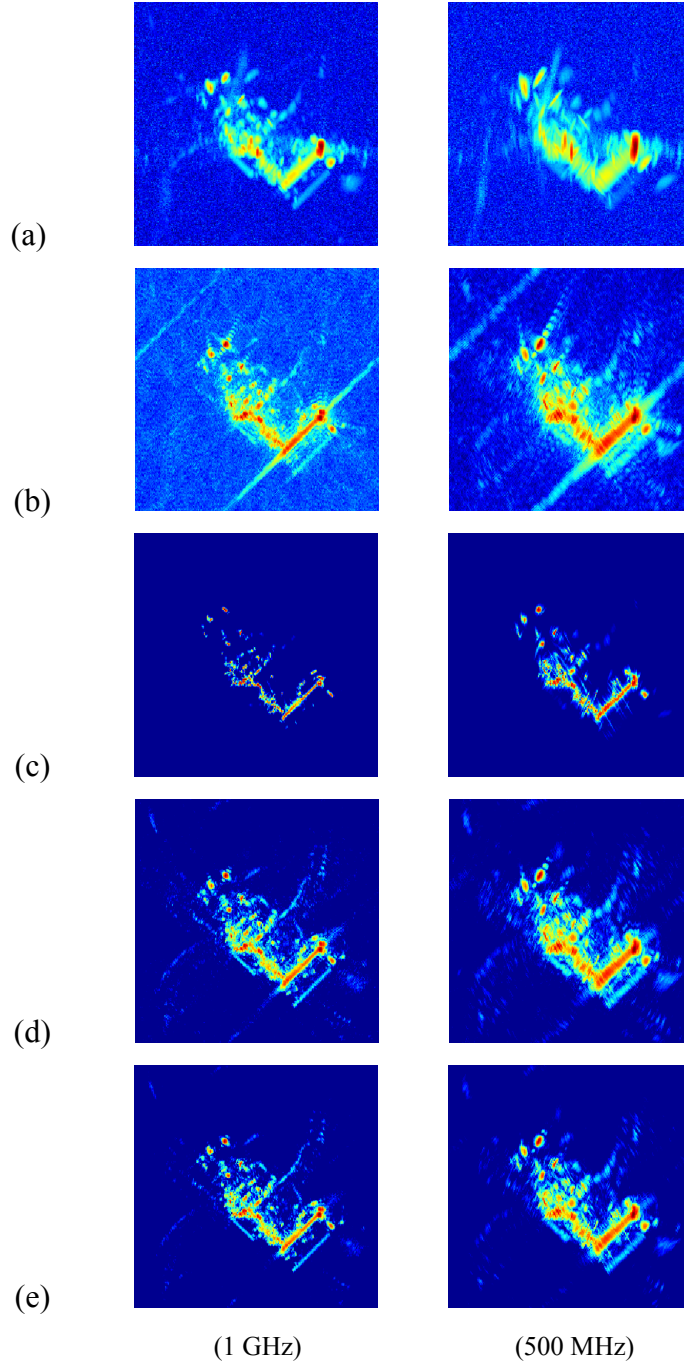
Figure 8. SAR images of the backhoe using bandwidths of 1 GHz and 500 MHz in the presence of SNR=20 dB. (a) Composite imaging. (b) Composite, point-enhanced imaging using small λ. (c) Composite, point-enhanced imaging using large λ. (d) Composite, point-enhanced imaging using λ selected by SURE. (e) Composite, point-enhanced imaging using λ selected by GCV.

parameters, respectively. Results from SURE and GCV are shown in (d) and (e), respectively. The conventional composite images do not preserve the scatterers of the backhoe. Small and large parameters have the effects mentioned before and displayed in Figure 4. SURE and GCV are able to choose an acceptable λ value in different noise levels and resolutions. Figure 7 and 8 also shows feature-enhanced composite images for different parameters obtained from data with SNR=20 dB and SNR=10 dB, respectively. Obviously, sensitivity to parameter choice increases at lower SNR's and SURE and GCV provide reasonable solutions.
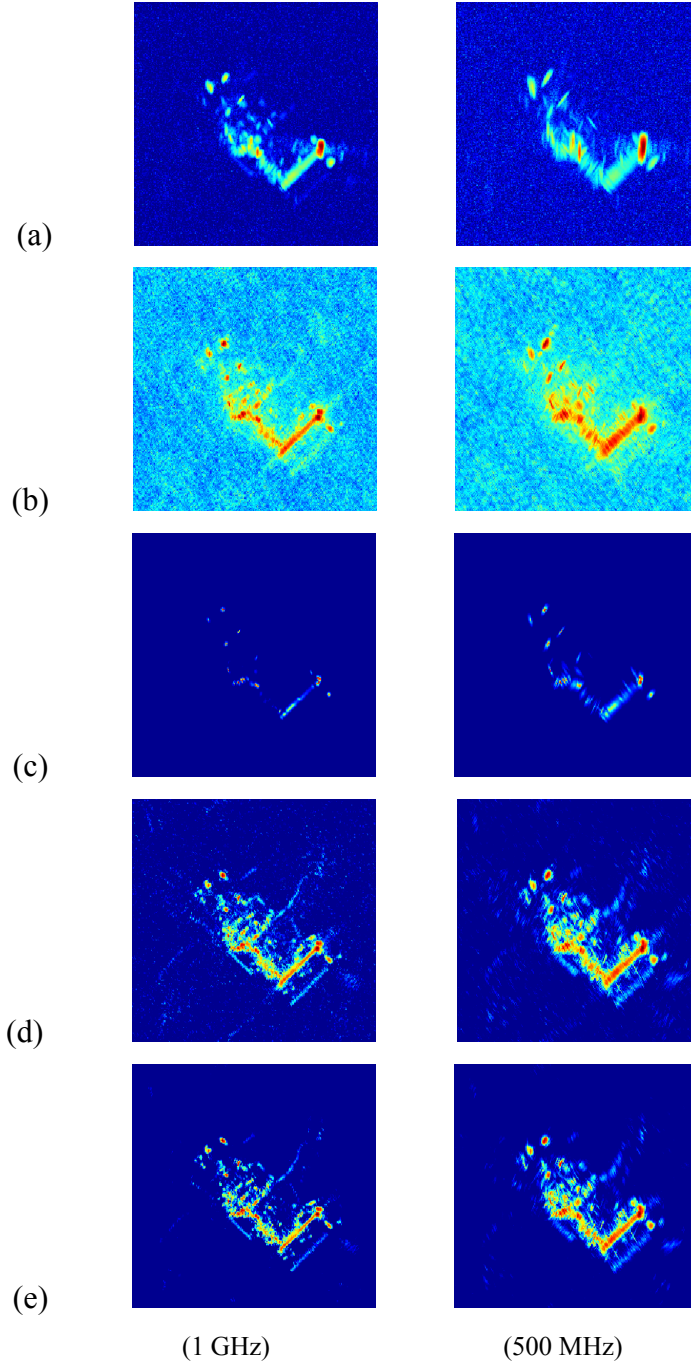
Figure 9. SAR images of the backhoe using bandwidths of 1 GHz and 500 MHz in the presence of SNR=10 dB. (a) Composite imaging. (b) Composite, point-enhanced imaging using small λ. (c) Composite, point-enhanced imaging using large λ. (d) Composite, point-enhanced imaging using λ selected by SURE. (e) Composite, point-enhanced imaging using λ selected by GCV.

6. CONCLUSION

We have considered the problem of hyper-parameter selection in non-quadratic regularization-based radar image formation. We have proposed to use SURE and GCV to select the regularization parameter for this problem and demonstrated images formed using parameters selected by SURE and GCV. We have proposed numerical solutions for

the optimization problems involved in these methods. We have observed that these methods lead to slight under-regularization but the parameter choices are reasonable. Regularized solutions become more sensitive to parameter choice at lower SNR's, thus the role of the parameter selection methods gains significance.

ACKNOWLEDGEMENTS

REFERENCES

[1] Çetin, M. and Karl, W. C., "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization, "IEEE Trans. on Image Processing 10(4), (2001).

[2] Benitz, G. R., "High-definition vector imaging", Lincoln Lab. J. 10, 147-170, (1997).

[3] DeGraaf, S. R., "SAR imaging via modern 2-D spectral estimation methods", *IEEE Trans. Image Processing*, 7, 729-761, (1998).

[4] Stein, M. S., "Estimation of the Mean of a Multivariate Normal Distribution", The Annals of Statistics, 9, 1135-1151, (1981).

[5] Golub, G. H., Heath, M. and Wahba, G., "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter", Technometrics, 21, 215-223, (1979).

[6] Solo, V., "A SURE-fired way to choose smoothing parameters in ill-posed inverse problems", Proc. IEEE ICIP96, IEEE, IEEE Press, (1996).

[7] "Backhoe data dome and Visual-D challange problem." Air Force Research Laboratory Sensor Data Management System (https://www.sdms.afrl.af.mil/main.php), (2004).

[8] Hutchinson, M. F., "A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines", Communications in Statistics, Simulation and Computation, 19, 433-450, (1990).

[9] Winston, W. L., and Venkataramanan, M., Introduction to Mathematical Programming, Brooks/Cole, USA, (2003).

[10] Nocedal, J. and Wright, S. J., Numerical Optimization, Springer, USA, (2006).

[11] Chen, S. S., Donoho, D. L. and Saunders, M. A., "Atomic Decomposition by Basis Pursuit", SIAM Journal of Scientific Computing, 20(1), 33-61, (1998).

[12] Moses, R. L., Potter, L. and Çetin, M., SPIE Defense and Security Symposium, Algorithms for Synthetic Aperture Radar Imagery XI, E. G. Zelnio and F. D. Garber, Eds., Orlando, Florida, (2004).

HOMOTOPY CONTINUATION FOR SPARSE SIGNAL REPRESENTATION

Dmitry M. Malioutov, Müjdat Çetin, and Alan S. Willsky

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
77 Massachusetts Ave., Cambridge, MA 02139, USA

ABSTRACT

We explore the application of a homotopy continuation-based method for sparse signal representation in overcomplete dictionaries. Our problem setup is based on the basis pursuit framework, which involves a convex optimization problem consisting of terms enforcing data fidelity and sparsity, balanced by a regularization parameter. Choosing a good regularization parameter in this framework is a challenging task. We describe a homotopy continuation-based algorithm to efficiently find and trace all solutions of basis pursuit as a function of the regularization parameter. In addition to providing an attractive alternative to existing optimization methods for solving the basis pursuit problem, this algorithm can also be used to provide an automatic choice for the regularization parameter, based on prior information about the desired number of non-zero components in the sparse representation. Our numerical examples demonstrate the effectiveness of this algorithm in accurately and efficiently generating entire solution paths for basis pursuit, as well as producing reasonable regularization parameter choices. Furthermore, exploring the resulting solution paths in various operating conditions reveals insights about the nature of basis pursuit solutions.

1. INTRODUCTION

Representing data in the most parsimonious fashion in terms of redundant collections of generating elements is at the core of many signal processing applications. However, finding such sparse representations exactly in terms of overcomplete dictionaries involves the solution of intractable combinatorial optimization problems. As a result, work in this area has focused on approximate methods, based on convex relaxations [1] or greedy methods, leading recently to the development of conditions under which such methods yield maximally sparse representations [2–6]. One such method, involving a convex ℓ_1 relaxation, is basis pursuit [1]. Its noisy version (allowing for some residual mismatch to data) poses the following optimization problem:

$$J(\mathbf{x}; \lambda) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1, \ \ \mathbf{A} \in \mathbb{R}^{M \times N} \quad (1)$$

where $\mathbf{y}$ denotes the data (signal whose representation we seek), $\mathbf{A}$ is the overcomplete representation dictionary ($M < N$), and $\lambda \geq 0$ is a scalar regularization parameter, balancing the tradeoff between sparsity and residual error. For a fixed λ, the problem can be solved by finding the minimizer $\hat{\mathbf{x}}$ of (1), using e.g. quadratic

programming. However choosing the regularization parameter is a difficult task, and some prior knowledge, either of the desired residual error (e.g. based on the noise level), or of the underlying sparse vector $\mathbf{x}$, has to be exploited. One piece of information about $\mathbf{x}$ might be the number of non-zero components. However, even if such information is available, how to use it directly in the basis pursuit framework is not straightforward.

Motivated by these observations, we describe a computationally efficient approach for sparse signal representation based on the homotopy continuation method of [7]. A related method has also been developed in [8], and has been linked to greedy methods. The main focus in [7] is the solution of an overdetermined least-squares problem with an ℓ_1-norm constraint. We are mostly interested in the unconstrained formulation in (1), in the underdetermined ($M < N$) case. In particular, we propose a simple algorithm to find and trace all solutions $\hat{\mathbf{x}}(\lambda)$ of basis pursuit as a function of the regularization parameter λ. The function $J(\mathbf{x}; \lambda)$ is convex and hence continuous, but it is not differentiable whenever $x_i = 0$ for some i, due to the term $\|\mathbf{x}\|_1 = \sum_i |x_i|$. The main idea of the approach is that $\|\mathbf{x}\|_1$, when restricted to the subset of non-zero indices of $\mathbf{x}$, is locally a linear function of $\mathbf{x}$. This allows one to solve the local problems (for a limited range of λ) analytically, and piece together local solutions to get solutions for all regions of λ. The resulting algorithm generates solutions for all λ with a computational cost that is comparable to solving basis pursuit with quadratic programming for a single λ. This procedure can also be used to select the regularization parameter λ based on information about the number of non-zero components in $\mathbf{x}$. In particular, a reasonable choice is the minimum λ that produces the desired number of non-zero components in $\hat{\mathbf{x}}(\lambda)$. Our numerical experiments demonstrate the effectiveness of this algorithm in generating the solution path accurately. Furthermore, exploring the structure of such solution paths reveals useful insights about the sensitivity of the problem to measurement noise, as well as to the nature of the overcomplete dictionary used.

2. NON-SMOOTH OPTIMALITY CONDITIONS

First we review non-smooth optimality conditions for convex functions and their implications for the problem in (1).

The subdifferential of a convex function $f : \mathbb{R}^N \to \mathbb{R}$ at $\mathbf{x} \in \mathbb{R}^N$ is defined as the following set:

$$\partial f(\mathbf{x}) = \{\xi \in \mathbb{R}^N | f(\mathbf{y}) \geq f(\mathbf{x}) + \xi^T(\mathbf{y} - \mathbf{x}) \ \forall \ \mathbf{y} \in \mathbb{R}^N\} \ (2)$$

Each element of $\partial f(\mathbf{x})$ is called a subgradient of f at $\mathbf{x}$. The subdifferential is a generalization of the gradient of f. In fact, if f

is convex *and* differentiable at a point $\mathbf{x}$ then

$$\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\} \qquad (3)$$

i.e. the subdifferential consists of a single vector, the gradient of f at $\mathbf{x}$ (the only subgradient is the gradient).

The non-smooth optimality conditions state that the subdifferential of f at $\mathbf{x}$ has to contain the $\mathbf{0}$-vector for f to achieve a global minimum at $\mathbf{x}$:

Theorem 1 (Non-smooth optimality conditions) *If $f : \mathbb{R}^N \to \mathbb{R}$ is convex, then f attains a global minimum at $\mathbf{x}$ if and only if $\mathbf{0} \in \partial f(\mathbf{x})$.*

The subdifferential of $g(\mathbf{x}) = \|\mathbf{x}\|_1$ is the following set:

$$\mathbf{u}(\mathbf{x}) \triangleq \partial g = \left\{ \mathbf{u} \in \mathbb{R}^N \;\middle|\; \begin{array}{ll} u_i = 1 & \text{if } x_i > 0 \\ u_i = -1 & \text{if } x_i < 0 \\ u_i \in [-1,..,1] & \text{if } x_i = 0 \end{array} \right\} \qquad (4)$$

The interesting part of this subdifferential is when some of the coordinates are equal to 0, where g is non-differentiable. Then u_i is not a scalar, it is a set.

The subdifferential of $f(\mathbf{x}) = J(\mathbf{x}; \lambda)$ from (1), for a fixed $\lambda = \tilde{\lambda}$, is the set

$$\partial f = \left\{ 2\mathbf{A}'(\mathbf{A}\mathbf{x} - \mathbf{y}) + \tilde{\lambda}\mathbf{u}(\mathbf{x}) \right\} \qquad (5)$$

where $\mathbf{u}(\mathbf{x})$ is defined above in (4). Suppose that $\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} J(\mathbf{x}; \tilde{\lambda})$. Then, in order to have $\mathbf{0} \in \partial f(\tilde{\mathbf{x}})$, the following equation must have a solution for some vector $\tilde{\mathbf{u}} \in \mathbf{u}(\tilde{\mathbf{x}})$:

$$2\mathbf{A}'\mathbf{A}\tilde{\mathbf{x}} + \tilde{\lambda}\tilde{\mathbf{u}} = 2\mathbf{A}'\mathbf{y} \qquad (6)$$

Let us consider an arbitrary vector $\mathbf{x}$ more closely. Let $\mathcal{I}_{on}$ be the support of $\mathbf{x}$, i.e. the set of indices i where $x_i \neq 0$. Also let $\mathcal{I}_{off}$ be the complement of $\mathcal{I}_{on}$, i.e. $\mathcal{I}_{off} = \{i \mid x_i = 0\}$. Put all entries x_i on the support of $\mathbf{x}$ into a vector $\mathbf{x}_{on}$, and the ones off the support of $\mathbf{x}$ into $\mathbf{x}_{off}$ (that makes $\mathbf{x}_{off} = \mathbf{0}$). Assume, without loss of generality, that $\mathbf{x}' = [\mathbf{x}'_{on} , \mathbf{x}'_{off}]$, i.e. the non-zero components appear first. Let us split $\mathbf{u}$ in the same fashion, according to which indices lie on or off the support of $\mathbf{x}$, into $\mathbf{u}_{on}$ and $\mathbf{u}_{off}$. Also, let us split the square $N \times N$ matrix $\mathbf{G} = 2\mathbf{A}'\mathbf{A}$ into 4 parts (there are 4 possibilities of whether the row-index and the column-index correspond to our sets $\mathcal{I}_{on}$ and $\mathcal{I}_{off}$): $\mathbf{G}_{on,on}$, $\mathbf{G}_{on,off}$, $\mathbf{G}_{off,on}$, $\mathbf{G}_{off,off}$. Due to symmetry of the matrix $\mathbf{G}$, we have $\mathbf{G}_{on,off} = \mathbf{G}'_{off,on}$. To simplify the notation further, let us use $\Phi = \mathbf{G}_{on,on}$, $\Psi = \mathbf{G}_{on,off}$, and $\Upsilon = \mathbf{G}_{off,off}$. Finally, let $\mathbf{z} = 2\mathbf{A}'\mathbf{y}$, and split $\mathbf{z}$ in the same way into $\mathbf{z}_{on}$ and $\mathbf{z}_{off}$.

Returning to our fixed $\tilde{\mathbf{x}}$ and $\tilde{\lambda}$, using our new notation, we can rewrite (6) as

$$\begin{pmatrix} \Phi & \Psi \\ \Psi' & \Upsilon \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{x}}_{on} \\ \mathbf{0} \end{pmatrix} + \tilde{\lambda} \begin{pmatrix} \tilde{\mathbf{u}}_{on} \\ \tilde{\mathbf{u}}_{off} \end{pmatrix} = \begin{pmatrix} \mathbf{z}_{on} \\ \mathbf{z}_{off} \end{pmatrix} \qquad (7)$$

Suppose that we know $\tilde{\mathbf{x}}$. The elements of $\tilde{\mathbf{u}}_{on}$ are all determined: they are equal to 1 or -1, corresponding to the signs of elements of $\tilde{\mathbf{x}}_{on}$. To determine $\tilde{\mathbf{u}}_{off}$, split equation (7) into two parts to get:

$$\Phi\tilde{\mathbf{x}}_{on} + \tilde{\lambda}\tilde{\mathbf{u}}_{on} = \mathbf{z}_{on} \qquad (8)$$

$$\Psi'\tilde{\mathbf{x}}_{on} + \tilde{\lambda}\tilde{\mathbf{u}}_{off} = \mathbf{z}_{off}$$

Thus we can find $\tilde{\mathbf{u}}_{off} = \frac{1}{\tilde{\lambda}}(\mathbf{z}_{off} - \Psi'\tilde{\mathbf{x}}_{on})$. Since $\tilde{\mathbf{x}}$ is optimal (for some $\lambda = \tilde{\lambda}$), the elements of $\tilde{\mathbf{u}}_{off}$ are constrained to lie in $[-1, 1]$.

3. FINDING SOLUTIONS FOR ALL λ

In the last section we characterized $\tilde{\mathbf{u}}$ given that we know $\tilde{\mathbf{x}}$, the optimal solution for a particular $\tilde{\lambda}$. Now starting with $\lambda = \tilde{\lambda}$, we incrementally change λ to find and trace optimal solutions $\hat{\mathbf{x}}(\lambda)$ for all λ. This forms the basis of the homotopy continuation method.

Suppose that $\tilde{\mathbf{x}}$ is the unique solution for $\tilde{\lambda}$ (where $\tilde{\lambda} > 0$), then from (8) we have[1]

$$\tilde{\mathbf{x}}_{on} = \Phi^{-1}(\mathbf{z}_{on} - \tilde{\lambda}\tilde{\mathbf{u}}_{on}) \qquad (9)$$

$$\tilde{\mathbf{u}}_{off} = \frac{1}{\tilde{\lambda}}(\mathbf{z}_{off} - \Psi'\Phi^{-1}\mathbf{z}_{on}) + \Psi'\Phi^{-1}\tilde{\mathbf{u}}_{on} \qquad (10)$$

No elements of $\tilde{\mathbf{x}}_{on}$ are equal to zero, hence there exists a range of λ, which includes $\tilde{\lambda}$, for which all entries of $\mathbf{x}_{on}(\lambda) = \Phi^{-1}(\mathbf{z}_{on} - \lambda\tilde{\mathbf{u}}_{on})$ will be nonzero. That means that throughout this range the support of $\mathbf{x}(\lambda)$ will not be reduced. By larger changes in λ we can force one of the components of $\mathbf{x}_{on}(\lambda)$ to zero. In addition, there exists a range of λ, which includes $\tilde{\lambda}$, for which $\mathbf{u}_{off}(\lambda) = \frac{1}{\lambda}(\mathbf{z}_{off} - \Psi'\Phi^{-1}\mathbf{z}_{on}) + \Psi'\Phi^{-1}\tilde{\mathbf{u}}_{on}$ does not become equal to 1 in absolute value, i.e. all entries of $\mathbf{u}_{off}(\lambda)$ belong to $[-1, 1]$. In the intersection of these two ranges of λ, the vectors $\mathbf{x}(\lambda)$ and $\mathbf{u}(\lambda)$ will satisfy the non-smooth optimality conditions for $J(\mathbf{x}(\lambda); \lambda)$, hence $\hat{\mathbf{x}}(\lambda) = \mathbf{x}(\lambda)$ for λ in the above region. The vector $\mathbf{x}(\lambda)$ is obtained by putting entries of $\mathbf{x}_{on}(\lambda)$ into the corresponding entries $\hat{x}_i(\lambda)$, for $i \in \mathcal{I}_{on}$, and zeros for $i \in \mathcal{I}_{off}$. The vector $\mathbf{u}(\lambda)$ is obtained by putting $\tilde{\mathbf{u}}_{on}$ (which does not change while λ is in the above region) into the components with $i \in \mathcal{I}_{on}$, and $\mathbf{u}_{off}(\lambda)$ for $i \in \mathcal{I}_{off}$.

In this way, we obtain all solutions for some range of λ's. The range can be easily calculated by solving for critical values of λ closest to $\tilde{\lambda}$, which make an entry of $\hat{\mathbf{x}}_{on}(\lambda)$ turn zero, or an entry of $\mathbf{u}_{off}(\lambda)$ reach unity in absolute value. This requires solving a set of scalar linear equations.

Now the next step is to find the support of $\hat{\mathbf{x}}(\lambda)$, as λ leaves the region. We only need to search locally, since $\hat{\mathbf{x}}(\lambda)$ is continuous for $\lambda > 0$ [7]. For the case where changing λ forces one component of $\mathbf{x}_{on}(\lambda)$ to zero, recalculating the support is trivial: we remove the index i for which x_i was set to zero from $\mathcal{I}_{on}$, and put it into $\mathcal{I}_{off}$. For the case where an entry of $\mathbf{u}_{off}(\lambda)$ becomes equal to 1 in absolute value, we transfer the corresponding index i from $\mathcal{I}_{off}$ into $\mathcal{I}_{on}$. The corresponding index of $\mathbf{u}_{on}$ is set to the sign of the entry of $\mathbf{u}_{off}(\lambda)$ which reached 1 in absolute value. Thus, after recomputing the support and the sign-pattern of solutions, we can proceed in the same fashion as before, computing the boundary of the new region for λ, finding the optimal solutions inside it, and entering a new region.

To start the algorithm, it is easiest[2] to consider $\lambda_0 = \infty$, or equivalently $\lambda_0 = 2\|\mathbf{A}'\mathbf{y}\|_\infty$, which satisfies $\hat{\mathbf{x}}(\lambda) = \mathbf{0}$ for $\lambda >$

---

[1]In this case, it can be shown that the matrix Φ is invertible.

[2]Another possibility is to start with $\lambda_0 = 0$, and increase it until $\hat{\mathbf{x}}(\lambda)$ becomes $\mathbf{0}$. Assuming that $\mathbf{A}$ has full row rank, this starting point requires the solution of the problem: $\min \|\mathbf{x}\|_1$ subject to $\mathbf{y} = \mathbf{A}\mathbf{x}$. The solution corresponds to $\lambda = 0^+$. When $\lambda = 0$ there exist multiple solutions if $\mathbf{A}$ has a nontrivial null-space. Solving the linear program picks the sparsest solution, which lies on the solution path $\hat{\mathbf{x}}(\lambda)$.

λ_0. Then, following the procedure described above, the algorithm produces $\hat{\mathbf{x}}(\lambda)$ for all $\lambda \geq 0$, and terminates when λ reaches 0.

The algorithm can exploit prior information about the desired number of non-zero elements in the representation to produce an automatic choice for the regularization parameter λ for basis pursuit. In particular, among all λ for which $\hat{\mathbf{x}}(\lambda)$ has the desired sparsity, the smallest one can be a reasonable choice in many scenarios, as it leads to the smallest residual, $\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}(\lambda)\|_2$. One might also consider other choices for λ, guided by the structure of the solution path, as we discuss in Section 4.

The computational complexity of the algorithm is dominated by the inversion of the matrix Φ at each breakpoint, which is bounded by $O(M^3)$, where M is the number of rows of $\mathbf{A}$. However, at each breakpoint the rank of the matrix Φ is changed by adding (or removing) a row and a column, hence instead of computing the inverse from scratch, rank-one updates can be done at the cost of $O(M^2)$. Empirically, the number of breakpoints is around M, but more careful analysis is in order. Thus, the cost of finding the whole solution path is roughly the same as for one iteration of the Newton's method to solve the problem in (1) for a fixed λ, i.e. $O(M^3)$. In addition, if one does not need the full solution path $\hat{\mathbf{x}}(\lambda)$, but only the path from $\hat{\mathbf{x}}(\lambda_0) = \mathbf{0}$ to a solution with L components, then the complexity is bounded by $O(L^3)$, with L instead of M, and the number of breakpoints is typically around L. Thus, the method is extremely efficient in computing very sparse solutions starting from $\hat{\mathbf{x}}(\lambda_0) = \mathbf{0}$.

To conclude the section, let us comment on the numerical stability of the algorithm. When we switch from one region to another, the only information that is carried over is the support of the new optimal solution, and the signs. Hence, if a small numerical error due to finite precision is made in computing the optimal solution for one region of λ (small enough not to affect the support and signs of the solution at the region boundary), then in the next region this error has no effect at all. Thus, the algorithm has a self-stabilizing property.

4. NUMERICAL EXAMPLES

4.1. Small Analytical Example

First we consider a very small example with $\mathbf{A} \in \mathbb{R}^{2 \times 3}$:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 1.5 \end{pmatrix}, \text{ and } \mathbf{y} = \begin{pmatrix} 6 \\ 6 \end{pmatrix}$$

We apply the algorithm from Section 3, and the resulting solution path is shown in Figure 1. For this small problem, we are also able to compute the entire solution path analytically, and observe that the algorithm produces it accurately. The two triangles are the intersections of $\mathbb{R}^+$ with the planes $x_1 + 2x_2 + 3x_3 = 6$, and $x_1 + 3x_2 + 1.5x_3 = 6$. The solution path $\hat{\mathbf{x}}(\lambda)$ starts at $\lambda = 60$, with $\mathbf{x} = \mathbf{0}$. As λ starts to decrease, the solution path enters a segment with one non-zero component: $x_2 = \frac{30}{13} - \frac{\lambda}{26}$, and $x_1 = x_3 = 0$. The segment satisfies optimality conditions until $\lambda = 28.8$, after which x_3 becomes non-zero. The solution path from $\lambda = 28.8$, down to $\lambda = 0^+$ is $x_2 = \frac{3}{2} - \frac{\lambda}{96}, x_3 = 1 - \frac{5\lambda}{144}$, and $x_1 = 0$. The minimum-norm solution, corresponding to $\lambda = 0$, is $\hat{\mathbf{x}}_{MN} = [.4968, 1.3758, .9172]$, is not sparse.

4.2. Larger Numerical Examples

Now we demonstrate the application of the algorithm on larger examples. We consider a problem $\mathbf{y} = \mathbf{Ax} + \mathbf{n}$, where $\mathbf{A}$ is an
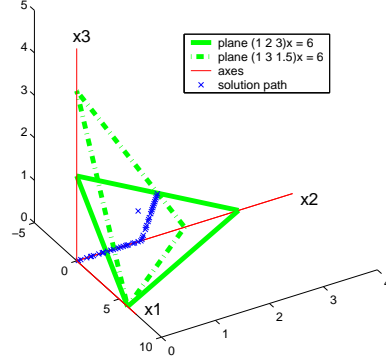

Fig. 1. Solution path for a small problem.

overcomplete 20×100 discrete cosine transform (DCT) dictionary, and $\mathbf{n}$ is zero-mean Gaussian noise. Dictionaries of this type arise naturally in many signal processing applications, one example being source localization with sensor arrays, where the observation model for linear arrays involves a discrete Fourier transform (DFT) dictionary [9]. In the specific example we consider here, $\mathbf{x}$ has two non-zero components, both equal to 1. In Figure 2 (top) we plot the solution path for noiseless data ($\mathbf{n} = \mathbf{0}$), in the middle plot for small amounts of noise (SNR = 15 dB), and in the bottom plot for moderate amounts of noise (SNR = 5 dB). Each piecewise-linear curve in these plots corresponds to one component $\hat{x}_i(\lambda)$. We also evaluate the solution at three intermediate values of λ in each linear segment, and compare it to a solution of the corresponding optimization problem in (1) using quadratic programming. The solutions agree almost perfectly, up to negligible numerical errors for all the examples.

Consider the top plot of Figure 2 which depicts the noiseless scenario. The smallest λ which leads to two non-zero components is $\lambda = 0^+$, which is the best parameter choice in this case. The corresponding solution found by homotopy-continuation has two non-zero entries equal to 1, and agrees with the original signal $\mathbf{x}$. In the middle plot, where the data are slightly noisy, the solution path ends at a non-sparse vector, which is close to the optimal solution of the noiseless problem (i.e. the other non-zero components are small). The smallest λ yielding exactly two non-zero components is $\lambda = 1.4548$. We note that the corresponding solution has non-zero indices not exactly equal, but very close to the ones of $\mathbf{x}$. The solution path suggests that an alternative to this choice of λ is to to pick a non-sparse solution for $\lambda = 0^+$ and threshold it, which would recover the exact indices in this mildly noisy scenario. In the bottom plot, the noise is sufficient to substantially change the solution path, but the smallest λ which leads to two non-zero elements ($\lambda = 0.6526$) still produces a reasonable solution, which is depicted in Figure 3 (we plot all components of $\hat{x}_i(\lambda)$ vs. i). Note that the indices of non-zero elements of $\hat{\mathbf{x}}(\lambda)$ are very close to those of the true $\mathbf{x}$. This 'stability' of indices of non-zero components occurs due to the special structure of $\mathbf{A}$: nearby columns of $\mathbf{A}$ are almost parallel for our overcomplete DCT matrix $\mathbf{A}$, and columns which are far apart are nearly orthogonal. This structure is what allows sparse signal representation ideas to be applied to source localization-type problems, even for highly overcomplete dictionaries [9].
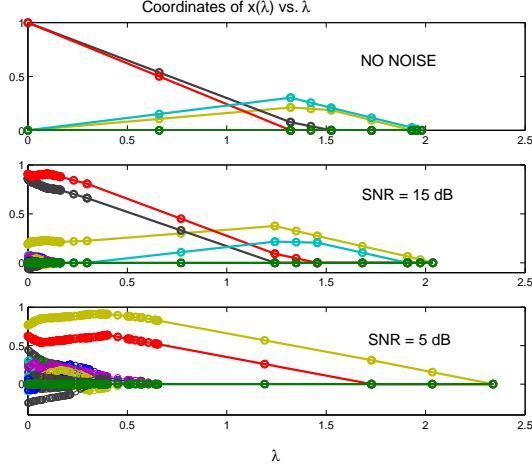
Fig. 2. Solution paths $\hat{\mathbf{x}}(\lambda)$ for all λ with varying levels of noise. $\mathbf{A}$ is 20×100. Top: no noise. Middle: SNR = 15 dB. Bottom: SNR = 5 dB.
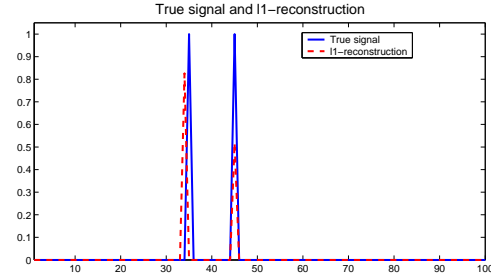

Fig. 3. $\hat{\mathbf{x}}(\lambda)$ for $\lambda = 0.6526$, the minimum λ leading to two non-zero components. SNR = 5 dB.

The above set of experiments were done for a severely overcomplete dictionary ($\mathbf{A}$ is 20×100). Let us now consider a mildly overcomplete, 20×23 DCT dictionary, $\mathbf{A}$. This problem is less demanding than the previous scenario in the sense that the desired signal representation is on a "coarser grid" of dictionary elements (leading to smaller mutual coherence [2]). In Figure 4, we observe that for noisy data the results exhibit excellent stability: even with moderate amounts of noise, SNR = 5 dB, the two non-zero components are clearly visible for any choice of λ. We note that these components exactly match the indices of non-zero elements of $\mathbf{x}$.

Some observations can be drawn from the above experiments. The components of $\hat{\mathbf{x}}(\lambda)$ tend to decrease as λ increases, but as can be seen from the middle plot in Figure 2, a component which was equal to 0 may become non-zero as λ increases. We also observe that sparse representation is easier in dictionaries with well-separated elements (in the sense of [2]). However, all hope is not lost even for severely overcomplete dictionaries, as long as they have certain structure.

5. CONCLUSION

We have described a simple and efficient algorithm to generate entire solution paths (as a function of the regularization parameter) of basis pursuit for sparse signal representation in overcomplete dic-
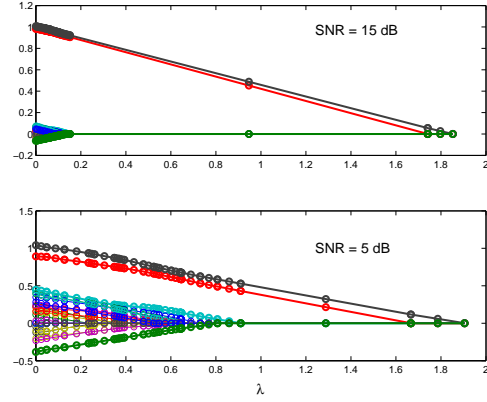

Fig. 4. Solution paths $\hat{\mathbf{x}}(\lambda)$ for all λ with varying levels of noise. $\mathbf{A}$ is 20×23. Top: SNR = 15 dB. Bottom: SNR = 5 dB.

tionaries. The algorithm can also be used to identify good choices for the regularization parameter. The ease in generating the solution paths make them a useful tool for empirical exploration of the behavior of basis pursuit in various scenarios.

6. REFERENCES

[1] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[2] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Aca. Sci.*, vol. 100, pp. 2197–2202, Mar. 2003.

[3] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Information Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[4] D. M. Malioutov, M. Çetin, and A. S. Willsky, "Optimal sparse representations in general overcomplete bases," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2004, vol. 2, pp. 793–796.

[5] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Information Theory*, vol. 50, no. 6, pp. 1341–1344, June 2004.

[6] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Information Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[7] M. R. Osborne, B. Presnell, and B. A. Turlach, "A new approach to variable selection in least squares problems," *IMA Journal of Numerical Analysis*, vol. 20, no. 3, pp. 389–403, 2000.

[8] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, Apr. 2004.

[9] D. M. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Processing*, to appear.

A Sparse Signal Reconstruction Perspective for Source Localization With Sensor Arrays

Dmitry Malioutov, *Student Member, IEEE*, Müjdat Çetin, *Member, IEEE*, and Alan S. Willsky, *Fellow, IEEE*

Abstract—We present a source localization method based on a sparse representation of sensor measurements with an overcomplete basis composed of samples from the array manifold. We enforce sparsity by imposing penalties based on the ℓ_1-norm. A number of recent theoretical results on sparsifying properties of ℓ_1 penalties justify this choice. Explicitly enforcing the sparsity of the representation is motivated by a desire to obtain a sharp estimate of the spatial spectrum that exhibits super-resolution. We propose to use the singular value decomposition (SVD) of the data matrix to summarize multiple time or frequency samples. Our formulation leads to an optimization problem, which we solve efficiently in a second-order cone (SOC) programming framework by an interior point implementation. We propose a grid refinement method to mitigate the effects of limiting estimates to a grid of spatial locations and introduce an automatic selection criterion for the regularization parameter involved in our approach. We demonstrate the effectiveness of the method on simulated data by plots of spatial spectra and by comparing the estimator variance to the Cramér–Rao bound (CRB). We observe that our approach has a number of advantages over other source localization techniques, including increased resolution, improved robustness to noise, limitations in data quantity, and correlation of the sources, as well as not requiring an accurate initialization.

Index Terms—Direction-of-arrival estimation, overcomplete representation, sensor array processing, source localization, sparse representation, superresolution.

I. INTRODUCTION

SOURCE localization using sensor arrays [1], [2] has been an active research area, playing a fundamental role in many applications involving electromagnetic, acoustic, and seismic sensing. An important goal for source localization methods is to be able to locate closely spaced sources in presence of considerable noise. Many advanced techniques for the localization of point sources achieve superresolution by exploiting the presence of *a small number* of sources. For example, the key component of the MUSIC method [3] is the assumption of a low-dimensional signal subspace. We follow a different approach for exploiting such a structure: We pose source localization as an overcomplete basis representation problem, where we impose a penalty on the lack of sparsity of the spatial spectrum.

Our approach is distinctly different from the existing source localization methods, although it shares some of their ingredients. The most well-known existing nonparametric methods

include beamforming [2], Capon's method [4], and subspace-based methods such as MUSIC [3]. Some additional methods (Root-MUSIC and ESPRIT) [1] require the assumption that the array of sensors is linear. Beamforming spectrum suffers from the Rayleigh resolution limit, which is independent of the SNR. MUSIC and Capon's method are able to resolve sources within a Rayleigh cell (i.e., achieve super-resolution), provided that the SNR is moderately high, the sources are not strongly correlated, and the number of snapshots is sufficient. A family of parametric methods based on the maximum likelihood paradigm, including deterministic maximum likelihood (DML) and stochastic maximum likelihood (SML) [1], enjoy excellent statistical properties, but an accurate initialization is required to converge to a global minimum. By turning to the sparse signal representation framework, we are able to achieve super-resolution without the need for a good initialization, without a large number of time samples, and with lower sensitivity to SNR and to correlation of the sources.

The topic of sparse signal representation has evolved very rapidly in the last decade, finding application in a variety of problems, including image reconstruction and restoration [5], wavelet denoising [6], feature selection in machine learning [7], radar imaging [8], and penalized regression [9]. There has also been some emerging investigation of these ideas in the context of spectrum estimation and array processing [10]–[14]. Sacchi *et al.* [10] use a Cauchy-prior to enforce sparsity in spectrum estimation and solve the resulting optimization problem by iterative methods. Jeffs [11] uses an ℓ_p-norm penalty with $p \leq 1$ to enforce sparsity for a number of applications, including sparse antenna array design. Gorodnitsky *et al.* [12] apply a recursive weighted minimum-norm algorithm called focal underdetermined system solver (FOCUSS) to achieve sparsity in the problem of source localization. It was later shown [15] that the algorithm is related to the optimization of ℓ_p penalties with $p \leq 1$. The work of Fuchs [13], [14] is concerned with source localization in the beamspace domain, under the assumption that the sources are uncorrelated, and that a large number of time samples is available. The method attempts to represent the vector of beamformer outputs to unknown sources as a sparse linear combination of vectors from a basis of beamformer outputs to isolated unit power sources. The method uses the ℓ_1 penalty for sparsity and the ℓ_2 penalty for noise. Prior research has established sparse signal representation as a valuable tool for signal processing, but its application to source localization has been developed only for very limited scenarios. We start with the ideas of enforcing sparsity by ℓ_1 penalties and extend them to a general framework that is applicable to a wide variety of practical source localization problems.

In its most basic form, the problem of sparse signal representation in overcomplete bases asks to find the sparsest signal

x to satisfy $\mathbf{y} = \mathbf{Ax}$, where $\mathbf{A} \in \mathbb{C}^{M \times N}$ is an overcomplete basis, i.e., $M < N$. Without the sparsity prior on $\mathbf{x}$, the problem $\mathbf{y} = \mathbf{Ax}$ is ill-posed and has infinitely many solutions. Additional information that $\mathbf{x}$ should be sufficiently sparse allows one to get rid of the ill-posedness. Solving problems involving sparsity typically requires combinatorial optimization, which is intractable even for modest data sizes; therefore, a number of relaxations have been considered [16]–[19]. We give a brief synopsis of relevant ideas in sparse signal representation in Section II.

The application of this methodology to practical array processing problems requires being able to handle additive noise, using multiple time or frequency samples from possibly strongly correlated sources in a sensible fashion, and allowing the data to be complex:

$$\mathbf{y}(t) = \mathbf{Ax}(t) + \mathbf{n}(t). \tag{1}$$

The goal of this paper is to explore how to utilize the sparse signal representation methodology for practical narrowband and wideband source localization using sensor arrays. The main contributions of our paper include a new adaptation of sparse signal representation to source localization through the development of an approach based on the singular value decomposition (SVD) to combine multiple samples and the use of second-order cone programming for optimization of the resulting objective function. The key ingredients of the proposed method is the use of SVD for data reduction and the formulation of a joint multiple-sample sparse representation problem in the signal subspace domain. In the body of the paper, we refer to the method as ℓ_1-SVD. In addition, we introduce the idea of adaptive grid refinement to combat the effects of a bias introduced by a limitation of the estimates to a grid. Finally, we discuss a method for the automatic selection of the regularization parameter involved in our approach, which balances data-fidelity with sparsity in the ℓ_1-SVD objective. In our experiments, the proposed approach exhibits a number of advantages over other source localization techniques, which include increased resolution, and improved robustness to noise, to limited number of snapshots, and to correlation of the sources. In addition, due to the convexity of all the optimization tasks involved in the approach, it does not require an accurate initialization. Another advantage of the approach is its flexibility, since few assumptions are made in the formulation, e.g., the array does not have to be linear, and the sources may be strongly correlated. Similarly, extensions to many scenarios, such as distributed sources and non-Gaussian noise, can be readily made. In the paper, we mostly focus on the narrowband farfield problem with arbitrary array geometry; we also describe the wideband scenario briefly in Section VIII-D. A more extensive discussion can be found in [20], where we also consider beamspace versions, cover wideband and nearfield processing in more detail, and propose an approach for simultaneous self-calibration and source localization in the presence of model errors.

We start with a brief introduction to the problem of sparse signal representation in Section II. In Section III, we describe the source localization problem and represent a single sample problem directly in the sparse signal representation framework.

In Section IV, we extend the approach to handle multiple samples. This is done in several steps, leading to the ℓ_1-SVD technique. In Section V, we describe how to find numerical solutions via a second-order cone programming (SOC) framework. We describe how to eliminate the effects of the grid in Section VI and propose how to automatically choose a regularization parameter involved in our approach in Section VII. Finally, in Section VIII, the advantages and disadvantages of the framework are explored using simulated experiments, and conclusions are made in Section IX.

II. SPARSE SIGNAL REPRESENTATION

The simplest version of the sparse representation problem without noise is to find a sparse $\mathbf{x} \in \mathbb{C}^N$, given $\mathbf{y} \in \mathbb{C}^M$, which are related by $\mathbf{y} = \mathbf{Ax}$, with $M < N$. The matrix $\mathbf{A}$ is known. The assumption of sparsity of $\mathbf{x}$ is crucial since the problem is ill-posed without it ($\mathbf{A}$ has a nontrivial null-space). An ideal measure of sparsity is the count of nonzero entries $\mathbf{x}$, which is denoted by $\|\mathbf{x}\|_0^0$, which we also call the ℓ_0-norm.[1] Hence, mathematically, we must look for $\arg \min \|\mathbf{x}\|_0^0$ such that $\mathbf{y} = \mathbf{Ax}$. This is, however, a difficult combinatorial optimization problem and is intractable for even moderately sized problems. Many approximations have been devised over the years, including greedy approximations (matching pursuit, stepwise regression, and their variants [17], [19]), as well as ℓ_1 and ℓ_p relaxations, where $\|\mathbf{x}\|_0^0$ is replaced by $\|\mathbf{x}\|_1$, [16], and $\|\mathbf{x}\|_p^p$, for $p < 1$, [20]. For the latter two, it has been shown recently that if $\mathbf{x}$ is "sparse enough" with respect to $\mathbf{A}$, then these approximations in fact lead to exact solutions (see [18], [20]–[24] for precise definitions of these notions).[2] In addition, [26] and [27] showed that with sufficient sparsity and a favorable structure of the overcomplete basis, sparse representations are stable in the presence of noise. These results are practically very significant since the ℓ_1 relaxation $\min \|\mathbf{x}\|_1$ subject to $\mathbf{y} = \mathbf{Ax}$ is a convex optimization problem, and the global optimum can be found for real-valued data by linear programming.[3] As these equivalence results are not specialized to the source localization problem but are derived for general overcomplete bases, the bounds that they provide are loose. A result that does take the structure of the basis into account is developed in [28].

In practice, a noiseless measurement model is rarely appropriate; therefore, noise must be introduced. A sparse representation problem with additive Gaussian noise takes the following form:

$$\mathbf{y} = \mathbf{Ax} + \mathbf{n}. \tag{2}$$

To extend ℓ_1-penalization to the noisy case, an appropriate choice of an optimization criterion is $\min \|\mathbf{x}\|_1$ subject to $\|\mathbf{y} - \mathbf{Ax}\|_2^2 \leq \beta^2$, where β is a parameter specifying how

---

[1] The symbols $\|\mathbf{x}\|_0$ and $\|\mathbf{x}\|_0^0$ are both used in the literature to represent the count of nonzero elements. We use the latter symbol since in the limit as $p \to 0^+$, $\|\mathbf{x}\|_p^p$ approaches the count of nonzero elements, but, if $\mathbf{x} \neq \mathbf{0}$ $\|\mathbf{x}\|_p \to \infty$.

[2] Recent studies of greedy methods, which have lower complexity than ℓ_1 and ℓ_p-based methods, have also yielded theoretical results of a similar flavor [25], [26].

[3] In addition, for the ℓ_p problem, local minima can be readily found by continuous optimization methods, as described in [20].

much noise we wish to allow. An unconstrained form of this objective is

$$\min \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1. \tag{3}$$

This objective function has been used in a number of sparse signal representation works ([16], [29] for real-valued data and [30] for complex-valued data). The ℓ_2-term forces the residual $\mathbf{y} - \mathbf{A}\mathbf{x}$ to be small, whereas the ℓ_1-term enforces sparsity of the representation. The parameter λ controls the tradeoff between the sparsity of the spectrum and the residual norm. We use these ideas in Sections III and IV for source localization.

The optimization criterion is again a convex optimization problem and can be readily handled by quadratic programming for real data. We propose the use of SOC programming for the complex data case. We describe SOC programming in Section V.

The class of methods called FOCUSS [12] is another paradigm for solving sparse signal representation problems with a more general ℓ_p penalty instead of ℓ_1. However, for $p < 1$, the cost function is nonconvex, and the convergence to global minima is not guaranteed. The discussion in [15] in Section VI indicates that the best results are obtained for p close to 1, whereas the convergence is also slowest for $p = 1$. The cost per iteration for FOCUSS methods is similar to that of an interior point solver for SOC since both solve a modified Newton's method step of similar dimensions. However, the number of iterations of SOC is better behaved (in fact, there are bounds on the worst-case number of iterations for SOC [31]) than for FOCUSS with $p = 1$. In our previous work [20], we have also observed slow convergence of iterative algorithms for ℓ_p minimization when applied with $p = 1$. By using an SOC formulation that is tailored to the convex ℓ_1 case, we are able to achieve fast convergence and guarantee global optimality of the solution.

III. SOURCE LOCALIZATION FRAMEWORK

A. Source Localization Problem

The goal of sensor array source localization is to find the locations of sources of wavefields that impinge on an array consisting of a number of sensors. The available information is the geometry of the array, the parameters of the medium where wavefields propagate, and the measurements on the sensors.

For purposes of exposition, we first focus on the narrowband scenario and delay the presentation of wideband source localization until Section VIII-D. Consider K narrowband signals $u_k(t), k \in \{1, \ldots, K\}$, arriving at an array of M omnidirectional sensors, after being corrupted by additive noise $n_m(t)$, resulting in sensor outputs $y_m(t), m \in \{1, \ldots, M\}$. Let $\mathbf{y}(t) = [y_1(t), \ldots, y_M(t)]'$ and similarly define $\mathbf{u}(t)$ and $\mathbf{n}(t)$. After demodulation, the basic narrowband observation model can be expressed as [1], [2]

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{u}(t) + \mathbf{n}(t), \quad t \in \{t_1, \ldots, t_T\}. \tag{4}$$

The matrix $\mathbf{A}(\boldsymbol{\theta})$ is the so-called array manifold matrix, whose (m, k)th element contains the delay and gain information from the kth source (at location θ_k) to the mth sensor. The columns

$\mathbf{a}(\theta_k)$ of $\mathbf{A}(\boldsymbol{\theta})$, for $k \in \{1, \ldots, K\}$, are called steering vectors. The number of sources K is unknown. To simplify the exposition, we only discuss the farfield scenario and confine the array to a plane, although neither of these assumptions is required for our approach. With farfield sources in the same plane as the array, the unknown locations of the sources are parameterized by angles (directions of arrival) with respect to the array axis $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_K]$. Given the knowledge of $\mathbf{y}(t)$ and the mapping $\boldsymbol{\theta} \to \mathbf{A}(\boldsymbol{\theta})$, the goal is to find the unknown locations of the sources θ_k for all k, as well as their number K.

B. Overcomplete Representation for a Single Time Sample

Now, we start to formulate the source localization problem as a sparse representation problem. The single-sample formulation in this section parallels the one in [12], where it was presented as one of applications of FOCUSS algorithm. In addition, the work in [13] and [14] is based on a similar philosophy of transforming a parameter estimation problem into sparse spectrum estimation, which we discuss later in this section.

We consider the single time sample case in this section, with $T = 1$ in (4). The problem as it appears in (4) is a nonlinear parameter estimation problem, where the goal is to find $\boldsymbol{\theta}$. Matrix $\mathbf{A}(\boldsymbol{\theta})$ depends on the unknown source locations $\boldsymbol{\theta}$, so it is not known.

To cast this problem as a sparse representation problem, we introduce an overcomplete representation $\mathbf{A}$ in terms of all possible source locations. Let $\{\tilde{\theta}_1, \ldots, \tilde{\theta}_{N_\theta}\}$ be a sampling grid of all source locations of interest. The number of potential source locations N_θ will typically be much greater than the number of sources K or even the number of sensors M. We construct a matrix composed of steering vectors corresponding to each potential source location as its columns: $\mathbf{A} = [\mathbf{a}(\tilde{\theta}_1), \mathbf{a}(\tilde{\theta}_2), \ldots, \mathbf{a}(\tilde{\theta}_{N_\theta})]$. In this framework $\mathbf{A}$ is known and does not depend on the actual source locations $\boldsymbol{\theta}$.

We represent the signal field by an $N_\theta \times 1$ vector $\mathbf{s}(t)$, where the nth element $s_n(t)$ is nonzero and equal to $u_k(t)$ if source k comes from $\tilde{\theta}_n$ for some k and zero otherwise. For a single time sample, the problem is reduced to

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{n}. \tag{5}$$

In effect, this overcomplete representation allows us to exchange the problem of parameter estimation of $\boldsymbol{\theta}$ for the problem of sparse spectrum estimation of $\mathbf{s}$. As in numerous nonparametric source localization techniques, the approach forms an estimate of the signal energy as a function of hypothesized source location, which ideally contains dominant peaks at the true source locations. The central assumption is that the sources can be viewed as point sources, and their number is small. With this assumption, the underlying spatial spectrum is sparse (i.e., $\mathbf{s}$ has only a few nonzero elements), and we can solve this inverse problem via regularizing it to favor sparse signal fields using the ℓ_1 methodology, as described in Section II. The appropriate objective function for the problem is

$$\min \|\mathbf{y} - \mathbf{A}\mathbf{s}\|_2^2 + \lambda\|\mathbf{s}\|_1. \tag{6}$$

We discuss how λ is chosen in Section VII, but for now, we assume that a good choice can be made. The data for the model
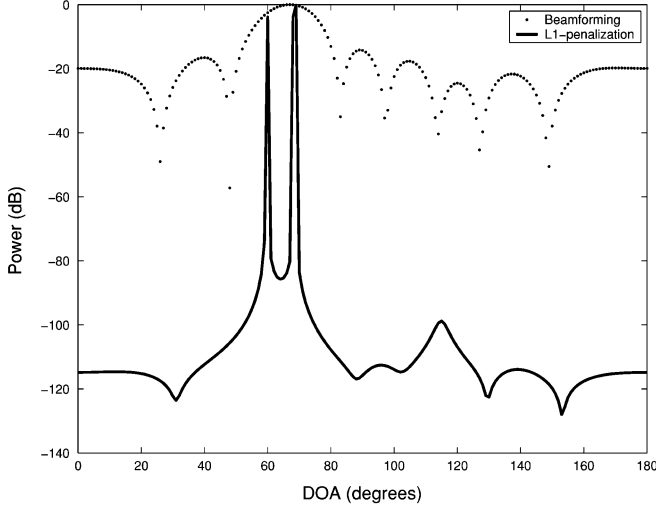
Fig. 1. Single sample source localization with ℓ_1. Spatial spectra of two sources with DOAs of 60° and 70° (SNR = 20 dB).

is complex-valued; hence, neither linear nor quadratic programming can be used for numerical optimization. Instead, we adopt an SOC programming framework, which we introduce in Section V. Once **s** is found, the estimates of the source locations correspond to the locations of the peaks in **s**.

We illustrate the approach for source localization with a *single time sample* in Fig. 1. We consider a uniform linear array of $M = 8$ sensors separated by half a wavelength of the actual narrowband source signals. We consider two narrowband signals in the far-field impinging on this array from directiions of arrival (DOAs) 60° and 70°, which are closer together than the Rayleigh limit. The SNR is 20 dB. The regularization parameter λ in this example is chosen by subjective assessment. We do not consider other source localization methods such as MUSIC or Capon's method in this simulation because they rely on estimating the covariance matrix of the sensor measurements, but in the simulation only, one time sample is present. Using beamforming, the two peaks of the spectrum are merged, but using the sparse regularization approach, they are well resolved, and the sidelobes are suppressed almost to zero. Apart from a small asymptotic bias, which we discuss in Section VIII, the spectrum estimate is an example of what super-resolution source localization methods aim to achieve.

The work of Fuchs [13], [14] is based on a similar philosophy of transforming a parameter estimation problem into a sparse spectrum estimation problem. A basis composed of beamformer outputs to isolated unit power sources from a large number of directions is created first. The method then attempts to represent the vector of beamformer outputs corresponding to the unknown sources as a sparse linear combination of vectors from the basis, using ℓ_1 penalties for sparsity, ℓ_2 penalties for noise, and optimization by quadratic programming. However, this beamspace domain formulation combines the multiple snapshots in a way that requires assumptions that the sources are uncorrelated and that a large number of samples is available. In contrast, the sensor-domain method that we propose in Section IV-C treats the multiple time samples in a very different way: We summarize multiple snapshots by using the SVD and solve a joint optimization problem over several singular vectors, imposing

a penalty that enforces the same sparsity profile over all these vectors, thus imposing temporal coherence. The resulting formulation is considerably more general than the one in [14].

IV. SOURCE LOCALIZATION WITH MULTIPLE TIME SAMPLES AND ℓ_1-SVD

Single snapshot processing may have its own applications, but source localization with multiple snapshots[4] from potentially correlated sources is of greater practical importance. When we bring time into the picture, the overcomplete representation is easily extended. The general narrowband source localization problem with multiple snapshots reformulated using an overcomplete representation has the following form:

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \quad t \in \{t_1, \dots t_T\}. \tag{7}$$

However, the numerical solution of this problem is a bit more involved than that of the single sample case. In Section IV-A, we describe a simple and computationally efficient method that, however, does not use the snapshots in synergy. In Section IV-B, we propose a coherent method that does use the snapshots in synergy but is more demanding computationally, and in Section IV-C, we develop an SVD-based approach that dramatically reduces the computational complexity while still using the snapshots coherently.

A. Treating Each Time Index Separately

The first thought that comes to mind when we switch from one time sample to several time samples is to solve each problem indexed by t separately. In that case, we would have a set of T solutions $\hat{\mathbf{s}}(t)$. If the sources are moving fast, then the evolution of $\hat{\mathbf{s}}(t)$ is of interest, and the approach is suitable for displaying it. However, when the sources are stationary over several time samples, then it is preferable to combine the independent estimates $\hat{\mathbf{s}}(t)$ to get one representative estimate of source locations from them, for example, by averaging or by clustering. This is noncoherent averaging, and its main attraction is its simplicity. However, by turning to fully coherent combined processing, as described in the following sections, we expect to achieve greater accuracy and robustness to noise.

B. Joint-Time Inverse Problem

Now, we consider a simple approach that uses different time samples in synergy. Let $\mathbf{Y} = [\mathbf{y}(t_1), \dots, \mathbf{y}(t_T)]$, and define $\mathbf{S}$ and $\mathbf{N}$ similarly. Then, (7) becomes

$$\mathbf{Y} = \mathbf{A}\mathbf{S} + \mathbf{N}. \tag{8}$$

There is an important difference of (8) from (5): Matrix $\mathbf{S}$ is parameterized temporally and spatially, but sparsity only has to be enforced in space since the signal $\mathbf{s}(t)$ in not generally sparse in time. To accommodate this issue, we impose a different prior: one that requires sparsity in the spatial dimension but does not require sparsity in time. This can be done[5] by first computing the ℓ_2-norm of all time-samples of a particular spatial index of **s**,

---

[4]While here we focus on multiple time snapshots, we will also use the same ideas applied to frequency snapshots for wideband source localization in Section VIII.

[5]It came to our attention that a similar idea has been used in [30] for basis selection.

i.e., $s_i^{(\ell_2)} = \|[s_i(t_1), s_i(t_2), \ldots, s_i(t_T)]\|_2$, and penalizing the ℓ_1-norm of $\mathbf{s}^{(\ell_2)} = [s_1^{(\ell_2)}, \ldots, s_{N_\theta}^{(\ell_2)}]$. The cost function becomes

$$\min \|\mathbf{Y} - \mathbf{AS}\|_f^2 + \lambda \|\mathbf{s}^{(\ell_2)}\|_1. \qquad (9)$$

The Frobenius norm is defined as $\|\mathbf{Y} - \mathbf{AS}\|_f^2 = \|\text{vec}(\mathbf{Y} - \mathbf{AS})\|_2^2$. The optimization is performed over $\mathbf{S}$; $\mathbf{s}^{(\ell_2)}$ is a function of $\mathbf{S}$. The time samples are combined using the 2-norm, which has no sparsifying effects. The spatial samples are combined using the ℓ_1-norm, which does enforce sparsity. Compared to the independent sample by sample processing from Section IV-A, the different time-indices of $\mathbf{s}$ reinforce each other, since the penalty is higher if the supports of $\mathbf{s}(t)$ for different t do not line up exactly. Once an estimate of $\mathbf{S}$ is computed using the new cost function, the peaks of $\mathbf{S}$ provide the source locations.

The main drawback of this technique is its computational cost. The size of the inverse problem increases linearly with T, and the computational effort required to solve it increases superlinearly with T. Thus, when T is large, this approach is not viable for the solution of the real-time source localization problem. We propose a solution to this problem next.

C. ℓ_1-SVD

In this section, we present a tractable approach to use a large number of time samples coherently, thus extending the use of sparse signal representation ideas for practical source localization problems. To reduce both the computational complexity and the sensitivity to noise, we use the SVD of the $M \times T$ data matrix $\mathbf{Y} = [\mathbf{y}(t_1), \ldots, \mathbf{y}(t_T)]$. The idea is to decompose the data matrix into the signal and noise subspaces, keep the signal subspace, and mold the problem with reduced dimensions into the multiple-sample sparse spectrum estimation problem in the form of Section IV-B. Note that we keep the signal subspace and not the noise subspace, which gets used in MUSIC, Pisarenko, and the minimum norm subspace methods.

Without noise on the sensors, the set of vectors $\{\mathbf{y}(t_i)\}_{i=1}^T$ would lie in a K-dimensional subspace, where K is the number of sources.[6] We would only need to keep a basis for the subspace (K vectors instead of T) to estimate what sparse combinations of columns of $\mathbf{A}$ form it. With additive noise, we decompose the data matrix into its signal and noise subspaces and keep a basis for the signal subspace. Mathematically, this translates into the following representation. Take the SVD[7]

$$\mathbf{Y} = \mathbf{ULV}'. \qquad (10)$$

Keep a reduced $M \times K$ dimensional matrix $\mathbf{Y}_{SV}$, which contains most of the signal power $\mathbf{Y}_{SV} = \mathbf{ULD}_K = \mathbf{YVD}_K$, where $\mathbf{D}_K = [\mathbf{I}_K \mathbf{0}']$. Here, $\mathbf{I}_K$ is a $K \times K$ identity matrix, and $\mathbf{0}$ is a $K \times (T-K)$ matrix of zeros. In addition, let $\mathbf{S}_{SV} = \mathbf{SVD}_K$, and $\mathbf{N}_{SV} = \mathbf{NVD}_K$, to obtain

$$\mathbf{Y}_{SV} = \mathbf{AS}_{SV} + \mathbf{N}_{SV}. \qquad (11)$$

[6]If $T < K$, or if the sources are coherent, we use the number of signal subspace singular values instead of K.

[7]This is closely related to the eigen-decomposition of the correlation matrix of the data: $\mathbf{R} = 1/T \mathbf{YY}'$. Its eigen-decomposition is $\mathbf{R} = 1/T \mathbf{ULV}'\mathbf{VL}'\mathbf{U}' = 1/T \mathbf{UL}^2\mathbf{U}'$.
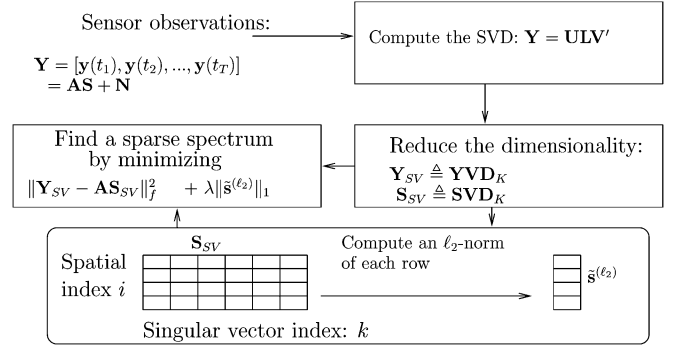

Fig. 2.   Block diagram of steps for ℓ_1-SVD.

Now, let us consider this equation column by column (each column corresponds to a signal subspace singular vector):

$$\mathbf{y}^{SV}(k) = \mathbf{As}^{SV}(k) + \mathbf{n}^{SV}(k), \quad k = 1, \ldots, K. \qquad (12)$$

This is now in exactly the same form as the original multiple time sample problem (7), except that instead of indexing samples by time, we index them by the singular vector number. What we have achieved by bringing the SVD transformation into the picture is the reduction of the size of the problem in Section IV-B from T blocks of data to K, where K is the number of sources. For typical situations where the number of sources is small and the number of time samples may be in the order of hundreds, this reduction in complexity is very substantial.

If we think of $\mathbf{S}_{SV}$ as a two-dimensional (2-D) field, indexed by i in the spatial dimension, and by k in terms of the singular vector index, then we again want to impose sparsity in $\mathbf{S}_{SV}$ only spatially (in terms of i) and not in terms of the singular vector index k. Similarly to Section IV-B, we define $\tilde{s}_i^{(\ell_2)} = \sqrt{\sum_{k=1}^K (s_i^{SV}(k))^2}, \forall i$. The sparsity of the resulting $N_\theta \times 1$ vector $\tilde{\mathbf{s}}^{(\ell_2)}$ corresponds to the sparsity of the spatial spectrum. We can find the spatial spectrum of $\tilde{\mathbf{s}}$ by minimizing

$$\|\mathbf{Y}_{SV} - \mathbf{AS}_{SV}\|_f^2 + \lambda \|\tilde{\mathbf{s}}^{(\ell_2)}\|_1. \qquad (13)$$

We illustrate the steps for the ℓ_1-SVD method in Fig. 2.

Note that our formulation uses information about the number of sources K. However, we empirically observe that incorrect determination of the number of sources in our framework has no catastrophic consequences (such as complete disappearance of some of the sources as may happen with MUSIC) since we are not relying on the structural assumptions of the orthogonality of the signal and noise subspaces. Underestimating or overestimating K manifests itself only in gradual degradation of performance. This is illustrated in Section VIII.

V. SOC REPRESENTATION OF THE ℓ_1-SVD PROBLEM

Now that we have an objective function in (13) to minimize, we would like to do so in an efficient manner. The objective contains a term $\|\tilde{\mathbf{s}}^{(\ell_2)}\|_1 = \sum_{i=1}^{N_\theta} \sqrt{\sum_{k=1}^K (s_i^{SV}(k))^2}$, which is neither linear nor quadratic. We turn to SOC programming [32], which deals with the so-called SOC constraints of the form $\mathbf{s} : \|s_1, \ldots, s_{n-1}\|_2 \leq s_n$, i.e., $\sqrt{\sum_{i=1}^{n-1}(s_i)^2} \leq s_n$. SOC programming is a suitable framework for optimizing functions

that contain SOC, convex quadratic, and linear terms. The main reason for considering SOC programming instead of generic nonlinear optimization for our problem is the availability of efficient interior point algorithms for the numerical solution of the former, e.g., [33]. In addition to efficient numerical solution, SOC programming has a substantial theoretical foundation as a special case of semidefinite programming and convex conic programming. See [32] for details; we describe in the Appendix how to manipulate the problem in (13) into the SOC programming form:

$$\min p + \lambda q$$
$$\text{subject to } \|(\mathbf{z}'_1, \ldots, \mathbf{z}'_K)\|_2^2 \leq p, \quad \text{and} \quad \mathbf{1}'\mathbf{r} \leq q$$
$$\text{where } \sqrt{\sum_{k=1}^{K} (s_i^{\text{SV}}(k))^2} \leq r_i, \quad \text{for } i = 1, \ldots, N_\theta$$
$$\text{and } \mathbf{z}_k = \mathbf{y}^{\text{SV}}(k) - \mathbf{A}\mathbf{s}^{\text{SV}}(k), \quad \text{for } k = 1, \ldots, K. \quad (14)$$

For the numerical solution of our SOC problem, we use a package for optimization over self-dual homogeneous cones (which includes direct products of the positive orthant-constraints, SOC constraints, and semidefinite cone constraints), called SeDuMi [33]. In terms of computational complexity, the interior point method relies on iterations of modified Newton's method. One of the main attractions of interior point methods is that the number of these iterations typically stays quite low, independent of the size of the problem. For optimizing the ℓ_1-SVD objective function in SOCP framework using an interior point implementation, the cost[8] is $O((K \times N_\theta)^3)$ with the observation that the number of iterations is empirically almost independent of the size of the problem [31] (a theoretical worst-case bound on the number of iterations is $O((K \times N_\theta)^{0.5})$ [31]). The computational complexity is higher than that of [14] since we have a joint optimization problem over K singular vectors, leading to an additional factor of K^3. It is also higher than the cost of MUSIC, where the main complexity is in the subspace decomposition of the covariance matrix, which is $O(M^3)$. However, the benefit that we get in return is generality. For reference, for a problem with three sources impinging upon an array with eight sensors and having $1°$ sampling of the spatial location of the sources (180 points on the grid), the time required for optimization using a Matlab implementation of the code on Linux on a computer with an 800-MHz Pentium 3 processor is roughly 5 sec, with around 20 iterations.

VI. MULTIRESOLUTION GRID REFINEMENT

Thus far, in our framework, the estimates of the source locations are confined to a grid. We cannot make the grid very fine uniformly since this would increase the computational complexity significantly. We explore the idea of adaptively refining the grid in order to achieve better precision. The idea is a very natural one: Instead of having a universally fine grid, we make the grid fine only around the regions where sources are present. This requires an approximate knowledge of the locations of the
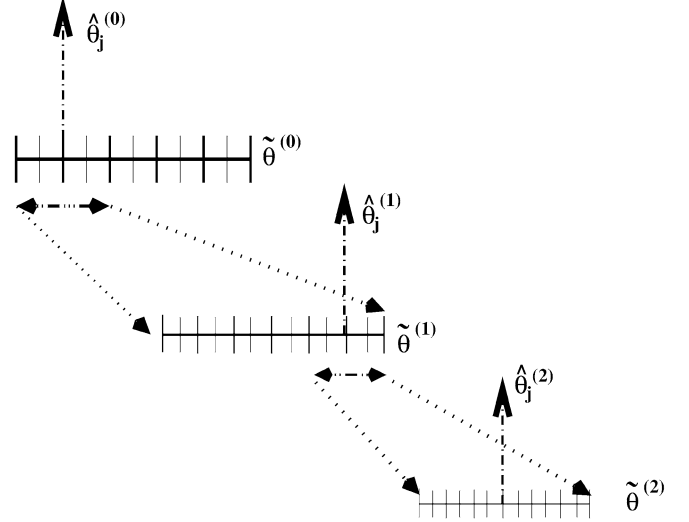
[8]We assume that $M \leq N_\theta$.

Fig. 3. Illustration of grid refinement.

sources, which can be obtained by using a coarse grid first. The algorithm is as follows.

1) Create a rough grid of potential source locations $\tilde{\theta}_i^{(0)}$, for $i = 1, \ldots, N_\theta$. Set $r = 0$. The grid should not be too rough in order to not introduce substantial bias. A $1°$ or $2°$ uniform sampling usually suffices.
2) Form $\mathbf{A}_r = \mathbf{A}(\tilde{\boldsymbol{\theta}}^{(r)})$, where $\tilde{\boldsymbol{\theta}}^{(r)} = [\tilde{\theta}_1^{(r)}, \tilde{\theta}_2^{(r)}, \ldots, \tilde{\theta}_{N_\theta}^{(r)}]$. Use our method from Section IV-C to get the estimates of the source locations $\hat{\theta}_j^{(r)}, j = 1, \ldots, K$, and set $r = r+1$.
3) Get a refined grid $\tilde{\boldsymbol{\theta}}^{(r)}$ around the locations of the peaks, $\hat{\theta}_j^{(r-1)}$. We specify how this is done below.
4) Return to step 2 until the grid is fine enough.

Many different ways to refine the grid can be imagined; we choose simple equispaced grid refinement. Suppose we have a locally uniform grid (piecewise uniform), and at step r, the spacing of the grid is δ_r. We pick an interval around the jth peak of the spectrum, which includes two grid spacings to either side, i.e., $[\hat{\theta}_j^{(r)} - 2\delta_r, \hat{\theta}_j^{(r)} + 2\delta_r]$, for $j = 1, \ldots, K$. In the intervals around the peaks, we select the new grid whose spacing is a fraction of the old one $\delta_{r+1} = \delta_r/\gamma$. It is possible to achieve fine grids either by rapidly shrinking δ_r for a few refinement levels or by shrinking it slowly using more refinement levels. We find that the latter approach is more stable numerically; therefore, we typically set $\gamma = 3$. After a few (e.g., 5) iterations of refining the grid, it becomes fine enough that its effects are negligible. Fig. 3 illustrates the refinement of the grid. The spacing of each of the grids corresponds to $2\delta_r$. The idea has been successfully used for some of the experimental analysis we present in Section VIII.

VII. REGULARIZATION PARAMETER SELECTION

An important part of our source localization framework is the choice of the regularization parameter λ in (13), which balances the fit of the solution to the data versus the sparsity prior. The same question arises in many practical inverse problems and is difficult to answer in many cases, especially if the objective function is not quadratic. We discuss an approach to select the regularization parameter automatically for the case where some

statistics of the noise are known or can be estimated. Let us denote the estimate of the spatial spectrum obtained using λ as the regularization parameter by $S(\lambda)$. A well-known idea under the name of discrepancy principle [34] is to select λ to match the residuals of the solution $\hat{S}(\lambda)$ to some known statistics of the noise when such are available. For example, if the distribution of the noise N_{SV} is known or can be modeled, then one can select λ such that $\|Y_{SV} - A\hat{S}(\lambda)\|_f^2 \approx E[\|N\|_f^2]$. Here, the we use the Frobenius norm $\|N\|_f^2 = \|\text{vec}(N)\|_2^2$. Directly searching for a value of λ to achieve the equality is rather difficult and requires solving the problem (13) multiple times for different λs.

Instead, we propose to look at the constrained version of the problem in (13), which can also be efficiently solved in the SOC framework [20]:

$$\min \|\tilde{s}^{(\ell_2)}\|_1 \text{ subject to } \|Y_{SV} - AS_{SV}\|_f^2 \le \beta^2. \quad (15)$$

The problem in (15) is equivalent via Lagrange multipliers to the one in (13) for some parameter β, which is related to λ. For the problem in (15), the task of choosing the regularization parameter β properly is considerably more transparent: We choose β high enough so that the probability that $\|\tilde{n}\|_2^2 \ge \beta^2$ is small, where $\tilde{n} = \text{vec}(NVD_K)$. If n is independent and identically distributed (i.i.d.) Gaussian, then for moderate to high SNR, $\|\tilde{n}\|_2^2$ has approximately a χ^2 distribution with MK degrees of freedom upon normalization by the variance of n. The reason that this holds only approximately is that the SVD in (10) $Y = AS + N = ULV'$ depends on the particular realization of noise, and hence, the matrix V is a function of N. However, when noise is small, the term AS dominates the SVD, and the change due to the addition of N is small, and we arrive at a χ^2 distribution for $\|\tilde{n}\|_2^2$. With the knowledge of the distribution, we can find a confidence interval for $\|\tilde{n}\|_2^2$ and use its upper value as a choice for β^2. In simulations we present in Section VIII, we find that this procedure generates appropriate regularization parameter choices for our problem when noise is reasonably small. We also present some thoughts on how to extend the range of the applicability of the procedure to higher levels of noise by characterizing the distribution of $\tilde{n}$ for lower SNR.

When noise statistics are not known, and no knowledge of the number of sources is available, the choice of the regularization parameter is a difficult question. It has been approached in the inverse problem community by methods such as L-curve [35]. An attempt to apply the L-curve to a subset selection problem in noise has been made in [36], but the authors have to make an assumption that the SNR is approximately known. The choice of the regularization parameter when no knowledge of the noise or of the sources is available is still an open problem.

VIII. EXPERIMENTAL RESULTS

In this section, we present several experimental results for our ℓ_1-SVD source localization scheme. First, we compare the spectra of ℓ_1-SVD to those of MUSIC [3], beamforming [2], Capon's method [4], and the beamspace method in [14] under various conditions. Next, we discuss and present results on regularization parameter selection. Then, we analyze empirically the bias and variance properties of our method. Finally, in Section VIII-D, we present an extension of our framework to
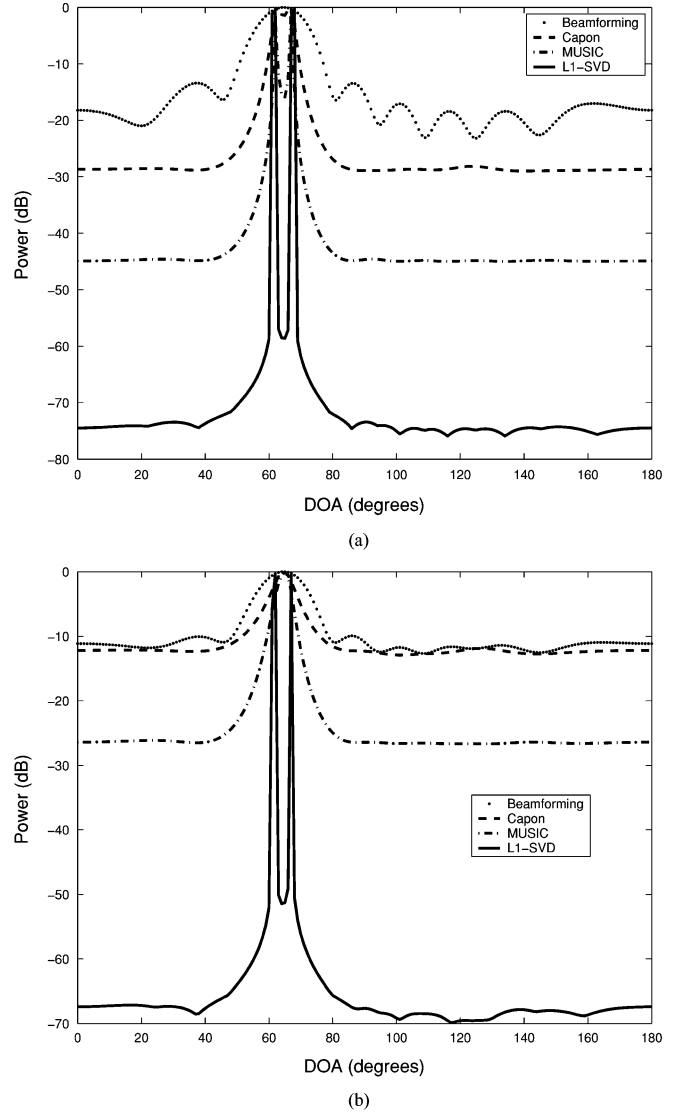

Fig. 4. (a) and (b). Spatial spectra for beamforming, Capon's method, MUSIC, and the proposed method (ℓ_1-SVD) for uncorrelated sources. DOAs: $62°$ and $67°$. Top: SNR $= 10$ dB. Bottom: SNR $= 0$ dB.

the wideband scenario and demonstrate its effectiveness on a number of examples.

A. Spectra for ℓ_1-SVD

We consider a uniform linear array of $M = 8$ sensors separated by half a wavelength of the actual narrowband source signals. Two zero-mean narrowband signals in the far-field impinge on this array from distinct DOAs. The total number of snapshots is $T = 200$, and the grid is uniform with $1°$ sampling $N_\theta = 180$. In Fig. 4, we compare the spectrum obtained using our proposed method with those of beamforming, Capon's method, and MUSIC. In the top plot, the SNR is 10 dB, and the sources are closely spaced ($5°$ separation). Our technique and MUSIC are able to resolve the two sources, whereas Capon's method and beamforming methods merge the two peaks. In the bottom plot, we decrease the SNR to 0 dB, and only our technique is still able to resolve the two sources. Next, we consider correlation between the sources, which can occur in practical array processing due to multipath effects. In Fig. 5, we set the SNR to 20
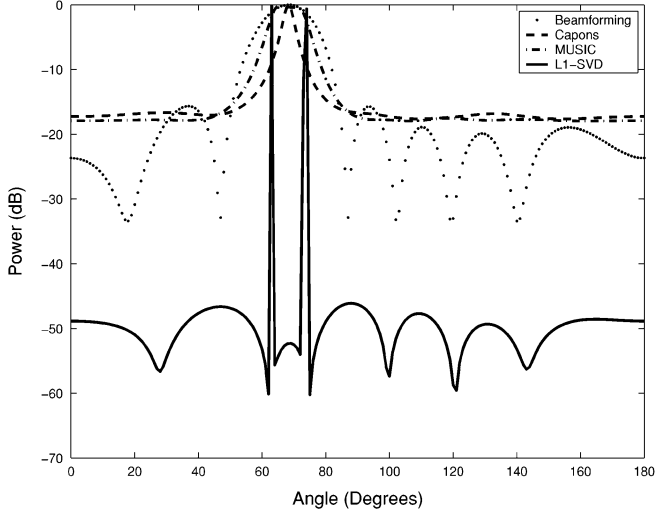
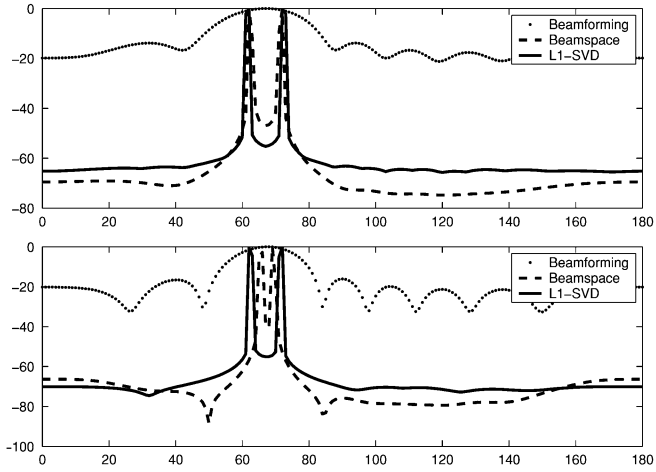Fig. 5. Spectra for correlated sources. SNR = 20 dB. DOAs: 63° and 73°.

Fig. 6. Comparison with beamspace technique of [14]. SNR = 20 dB. DOAs: 63° and 73°. Top: Uncorrelated sources. Bottom: Correlated sources; correlation coefficient is 0.99.

dB but make the sources strongly correlated, with a correlation coefficient of 0.99. MUSIC and Capon's method would resolve the sources at this SNR were they not correlated, but correlation degrades their performance. Again, only our technique is able to resolve the two sources. This illustrates the power of our methodology in resolving closely spaced sources despite low SNR or correlation between the sources.

In Fig. 6, we compare the spectra obtained using ℓ_1-SVD to spectra obtained using our implementation of the beamspace technique described in [14]. The top plot considers two uncorrelated sources at 63° and 73°, with $T = 200$ samples. SNR is 0 dB. As can be seen from the plot, for uncorrelated sources with $T = 200$, the assumptions made in [14] hold, and the beamspace method has an excellent performance, similar to that of our ℓ_1-SVD method.

In the bottom plot, the two sources are correlated, breaking the assumption in [14]. We observe that the performance of the beamspace technique degrades and that strong bias appears. This bias was not present when the sources were uncorrelated. As we already noted, no such degradation appears for ℓ_1-SVD,
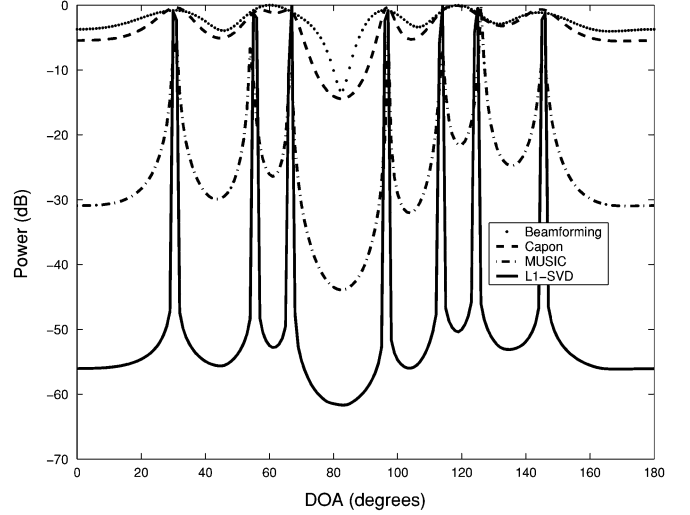

Fig. 7. Resolving $M - 1$ sources: $M = 8$ sensors, seven sources, SNR = 10 dB.

and the spectrum is very similar to the one for the case of uncorrelated sources. In summary, our formulation is based on similar principles of enforcing sparsity as the work in [14], but it is more general in allowing correlated sources and making no assumptions of having a large number of time samples.

Thus far, we have shown plots resolving a small number of sources. An interesting question is to characterize the maximum number of sources that can be resolved by ℓ_1-SVD using measurements from an M-sensor array. It can be shown through simple linear algebraic arguments that M sources cannot be localized (the representation is ambiguous). However, empirically, the ℓ_1-SVD technique can resolve $M - 1$ sources[9] if they are not located too close together. Hence, ℓ_1-SVD is not limited to extremely sparse spectra but can resolve the same number of sources as MUSIC and Capon's methods. This is illustrated in Fig. 7. The number of sensors in the array is again $M = 8$, and the number of sources is 7. With moderate SNR as in this example, all three techniques (ℓ_1-SVD, MUSIC, and Capon's method) exhibit peaks at the source locations.

We mentioned in Section IV-C that our approach is not very sensitive to the correct determination of the number of sources. We give an illustration of this statement in Figs. 8 and 9. We use the same $M = 8$ sensor uniform linear array as before. The actual number of sources is $K = 4$, and the SNR is 10 dB. In Fig. 8, we plot unnormalized (i.e., the maximum peak is not set to 1) spectra obtained using MUSIC when we vary the assumed number of sources. Underestimating the number of sources results in a strong deterioration of the quality of the spectra, including widening and possible disappearance of some of the peaks. A large overestimate of the number of sources leads to the appearance of spurious peaks due to noise. In Fig. 9, we plot the unnormalized spectra obtained using ℓ_1-SVD for the same assumed numbers of sources, and the variation in the spectra is

[9]This holds under the assumption that the number of singular vectors used in ℓ_1-SVD is sufficient, e.g., equal to the number of sources. When fewer singular vectors are taken than the number of sources, the number of resolvable sources may decrease. However, even in the extreme case of taking just one singular vector, for the eight-sensor array in the example in Fig. 9, ℓ_1-SVD resolves 4 (i.e., $M/2$) sources.
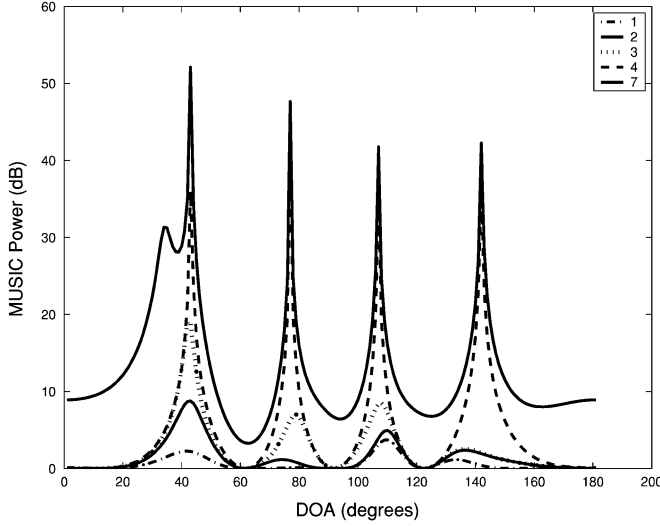
Fig. 8.   Sensitivity of MUSIC to the assumed number of sources. The correct number is 4.

Fig. 10.   Regularization parameter choice: Discrepancy principle leads to a useful spectrum. Setting the regularization parameter too low produces spurious peaks in the spectrum.
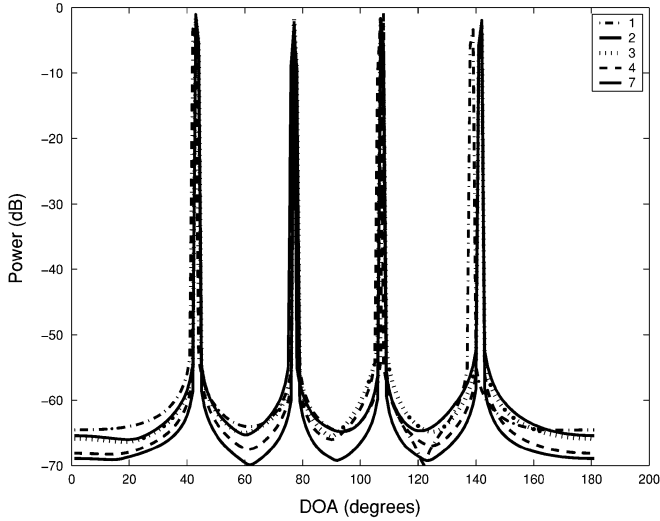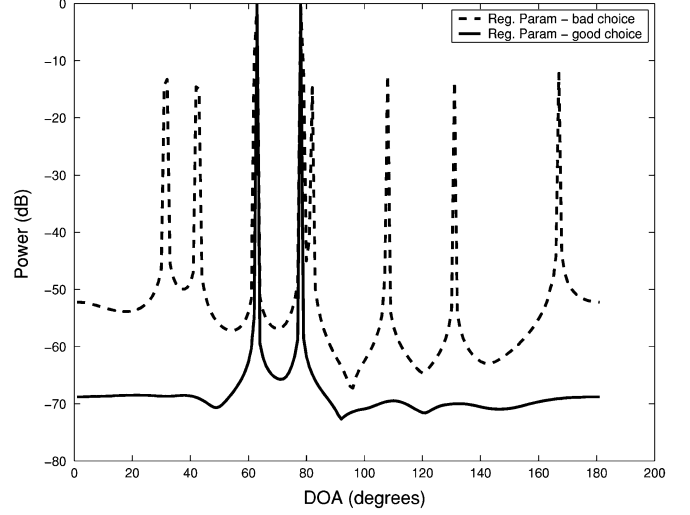

Fig. 9.   Sensitivity of ℓ_1-SVD to the assumed number of sources. The correct number is 4.

very small. The importance of the low sensitivity of our technique to the assumed number of sources is twofold. First, the number of sources is usually unknown, and low sensitivity provides robustness against mistakes in estimating the number of sources. In addition, even if the number of sources is known, low sensitivity may allow one to reduce the computational complexity of ℓ_1-SVD by taking a smaller number of singular vectors. With higher levels of noise, in our experiments, we observe that the sensitivity of ℓ_1-SVD to the assumed number of sources increases; however, it still provides better robustness relative to MUSIC, especially when the assumed number of sources is less than the actual number of sources.

B. Regularization Parameter Choice

We illustrate the importance of a good choice of the regularization parameter in Fig. 10. The number of sources in the example is $K = 2$, and the number of sensors and snapshots is kept as before $M = 8$ and $T = 200$. The curve labeled "good

choice" represents the selection of the regularization parameter β by the discrepancy principle from Section VII, with a 99% confidence interval. The spectrum is sharp, and the peaks correspond to source locations. For the second curve labeled "bad choice," the regularization parameter was set three times lower: below the norm of the realization of the noise. In order to explain the data with such a small regularization parameter, spurious peaks due to noise appear in the plot. In addition, if we set the regularization parameter too high, starting from about five times the value selected by the discrepancy principle one of the peaks would disappear, and as we increase it further, the second peak would disappear, making the spectrum 0 at all spatial locations. This example illustrates two points: the importance of a good choice of the regularization parameter and the soundness of the approach based on the discrepancy principle.

In Section VII, in order to calculate the confidence intervals for $\|\tilde{\mathbf{n}}\|_2^2$, we had to make an assumption that noise is reasonably small. When the assumption does not hold, the SVD $\mathbf{Y} = \mathbf{AS} + \mathbf{N} = \mathbf{ULV}'$ depends on $\mathbf{N}$, and $\mathbf{N}_{SV} = \mathbf{NVD}_K$ is a complicated function of $\mathbf{N}$ since $\mathbf{V}$ now depends on $\mathbf{N}$. One approach to characterize $\|\check{\mathbf{n}}\|_2^2$ for higher levels of noise is through simulation. In Fig. 11, we illustrate the dependence of the ratio of $\|\tilde{\mathbf{n}}\|_2/\sigma$ on SNR, where σ^2 is the variance of the i.i.d. Gaussian noise $\mathbf{n}(t)$. To create the plot, we first selected $K = 3$ source locations uniformly distributed in $[0, \pi]$, $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3]$, and a corresponding signal matrix $\mathbf{S}$, with indices of nonzero rows corresponding to $\boldsymbol{\theta}$. For each choice of $\boldsymbol{\theta}$, we created 250 instances of zero-mean i.i.d. Gaussian noise matrices $\mathbf{N}$ with variance σ^2 and calculated the minimum, average, and maximum ratios $\|\tilde{\mathbf{n}}\|_2/\sigma$ over all 250 instances of $\mathbf{N}$. The three curves (max, min, and average ratio) are plotted as a function of SNR. We superimposed these curves for ten different realizations of $\boldsymbol{\theta}$ to show the variability. For very low SNR, noise is dominating $\mathbf{Y}$:$\mathbf{Y} = \mathbf{AS} + \mathbf{N} \approx \mathbf{N}$ and $\|\tilde{\mathbf{n}}\|_2^2 \approx \sum_{k=1}^K \sigma_k^2$, where $\{\sigma_k\}_{k=1}^K$ are the top K singular values. For high SNR, noise has a small contribution to $\mathbf{Y}$, and $\|\tilde{\mathbf{n}}\|_2^2$ can be well predicted as described in Section VII. However, there is a sharp
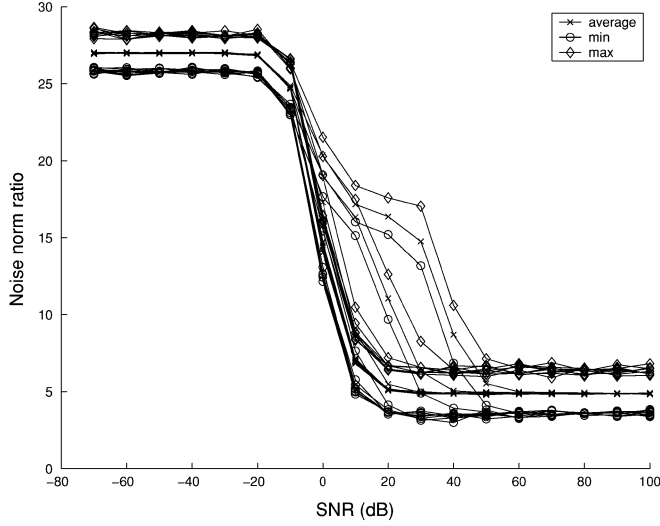
Fig. 11. Regularization parameter choice for moderate noise: Ratio of $\|\bar{\mathbf{n}}\|_2$ to σ as a function of SNR.

Fig. 12. Bias of ℓ_1-SVD in localizing two sources as a function of separation between the two sources SNR $= 10$ dB.

transition between these two regions, which we are interested in characterizing. For most triples of curves, the transition occurs at the same SNR, but there are two outliers. They occur when source locations $\boldsymbol{\theta}$ are closely spaced so that $\mathbf{A}(\boldsymbol{\theta})$ has a high condition number (recall that $\mathbf{A}(\boldsymbol{\theta})$ contains columns of $\mathbf{A}$ corresponding to $\boldsymbol{\theta}$). In that case, the effects of noise start to show up at higher SNR. The conclusion that can be drawn out of these experiments is that it is possible to predict $\|\tilde{\mathbf{n}}\|_2^2$ for higher levels of noise, but one has to be careful with closely spaced sources.

C. Bias and Variance

One aspect of our technique is the bias of the estimates that appears for closely spaced sources. The reason for the bias is that we impose a sparsity prior in our objective function, without which the problem of estimating the spectrum is ill-posed. Other source localization methods have much difficulty resolving closely spaced sources, especially at low SNRs; hence, small bias can be considered as a good compromise, if such peaks can be resolved. We now investigate bias[10] more closely by considering source localization with two sources and varying the angular separation between them. The number of sensors and snapshots is again $M = 8$ and $T = 200$. In Fig. 12, we plot the bias of each of the two source location estimates as a function of the angular separation when one source is held fixed at $42°$. The SNR is 10 dB. The values on each curve are an average over 50 trials. The plot shows the presence of bias for low separations, but the bias disappears when sources are more than about $20°$ apart.

We next compare the variance of the DOA estimates produced by our approach to those obtained using existing methods [1] and to the CRB. In order to satisfy the assumptions of the CRB, we choose an operating point where our method is unbiased, i.e., when the sources are not very close together. In Fig. 13, we present plots of variance versus SNR for a scenario
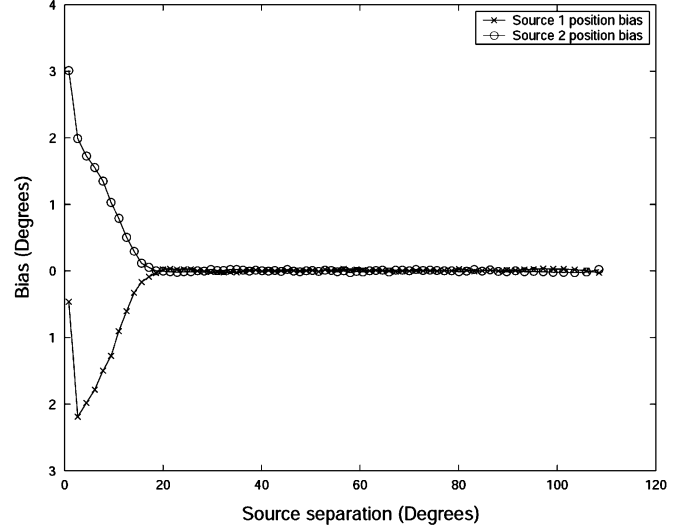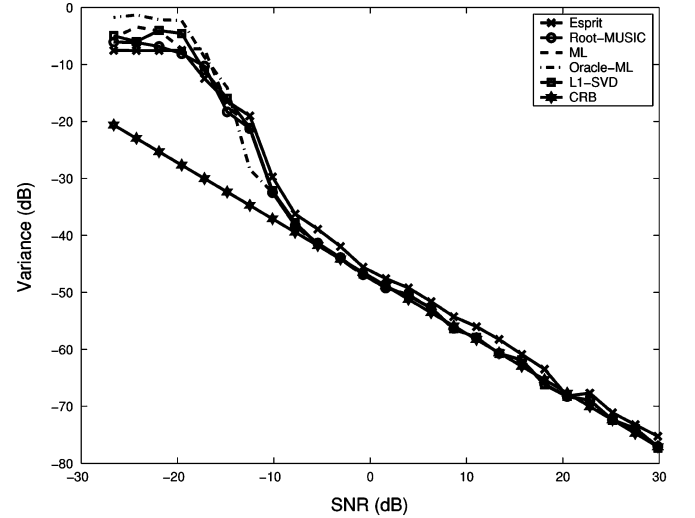

Fig. 13. CRB for zero mean *uncorrelated* sources. Comparison with variances of ESPRIT, Root-MUSIC, ML, and ℓ_1-SVD. DOAs: $42.83°$ and $73.33°$.

including two uncorrelated sources.[11] On the plot, we also include a curve labeled "oracle" maximum likelihood, which is obtained by using an ML estimate, where the nonconvex optimization is initialized to the true values of the source locations. This estimator is not practically realizable and intuitively serves as an effective bound for performance in the threshold region, where the CRB is rather loose. Each point in the plot is the average of 50 trials. It can be seen that for well-separated sources, the variance of ℓ_1-SVD estimates follows closely that of other estimators and, except for very low SNR, meets the CRB. As we have illustrated in Fig. 4, closely spaced sources can be resolved at lower SNR with our technique than is possible with other methods. This occurs in a region where our method is biased. On the other hand, Fig. 13 shows that when the sources are well-separated, and our method is unbiased, its performance is as good as

---

[10]Our analysis of bias and variance is based on computer simulations. The work in [29] contains a theoretical analysis of bias and variance in a limited scenario for one time sample and for a single source.

[11]To obtain this plot, we have used the adaptive grid refinement approach from Section VI to get point estimates not limited to a coarse grid.
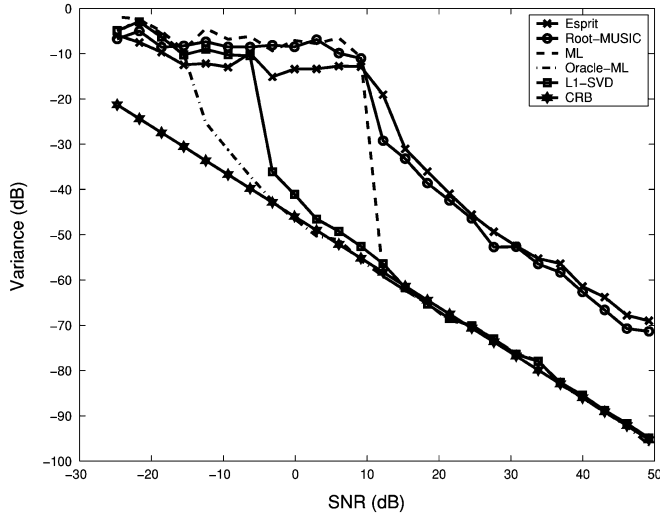
Fig. 14. Plots of variances of DOA estimates versus SNR, as well as the CRB, for two correlated sources. DOAs: $42.83°$ and $73.33°$. Variance for the source at $42.83°$ shown.

those of existing super-resolution methods. Another important advantage can be seen in Fig. 14 for correlated sources, which commonly occur in practice due to multipath effects. The correlation coefficient is 0.99. Our approach follows the CRB more closely than the other methods, and the threshold region occurs at lower SNR. The proposed ℓ_1-SVD method is the closest one in performance to the intuitive bound provided by the oracle-ML curve. This shows the robustness of our method to correlated sources.

D. Wideband Source Localization

The main difficulty that arises when wideband signals are considered is the impossibility to represent the delays by simple phase shifts. A way to deal with this issue is to separate the signal spectrum into several narrowband regions, each of which yields to narrowband processing. To work in the frequency domain, the time-samples are grouped into several "snapshots," and transformed into the frequency domain:

$$\mathbf{y}^{(n)}(\omega) = \mathbf{A}(\omega)\mathbf{s}^{(n)}(\omega) + \mathbf{n}^{(n)}(\omega), \quad n \in \{1, \ldots, N_s\}. \quad (16)$$

For each frequency ω, we have N_s snapshots. We are in general interested in a 2-D power spectrum as a function of both spatial location (DOA) and frequency ω; therefore, we solve the problem at each frequency independently, using the ℓ_1-SVD method, with frequency snapshots replacing temporal snapshots.

In Fig. 15, we present an example using the same eight-element uniform linear array as the one used throughout the paper, but the signals are now wideband. We consider three chirps with DOAs $70°$, $98°$, and $120°$ with frequency span from 250 to 500 Hz, and $T = 500$ time samples. Using conventional beamforming, the spatio-frequency spectra of the chirps are merged and cannot be easily separated [plot (a)], especially in lower frequency ranges, whereas using ℓ_1-SVD [plot (b)], they can be easily distinguished throughout their support. This shows that the ℓ_1-SVD methodology is useful for wideband scenarios as well.
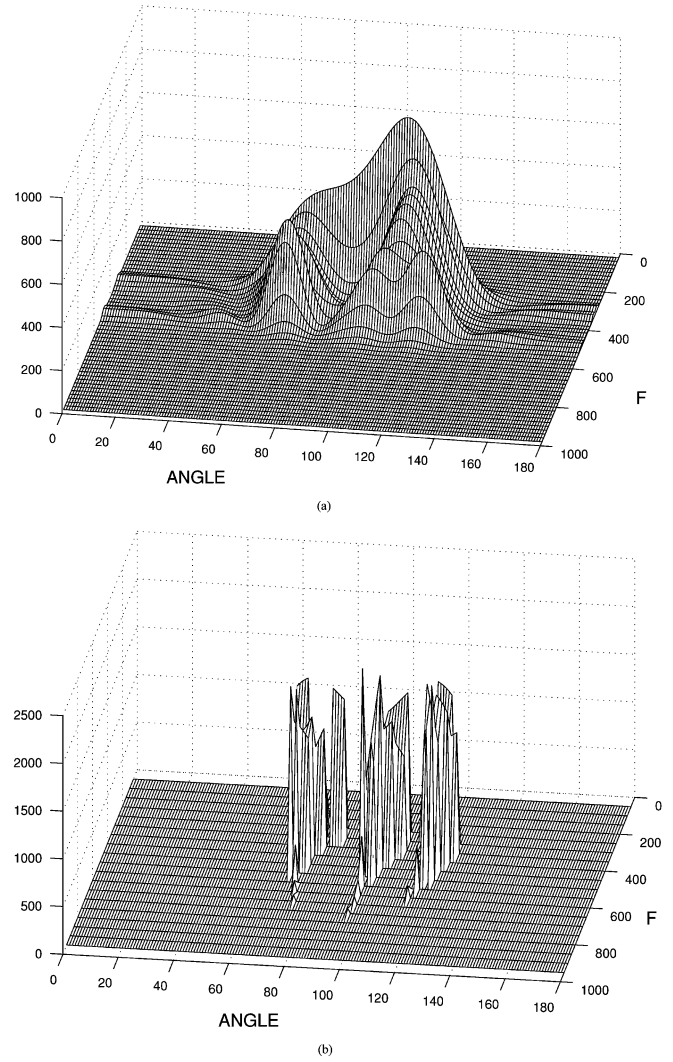

Fig. 15. (a) and (b). Wideband example: Three chirps. DOAs: $70°$, $98°$, and $120°$. Frequencies are processed independently. Top: Conventional beamforming. Bottom: ℓ_1-SVD processing.

The approach that we just described treats each frequency independently. In [20], we outline an alternative version of wideband source localization for joint "coherent" processing of the data at all frequencies. Wideband adaptations of current source localization methods, based on ideas such as focusing matrices [37], can do coherent processing over a narrow frequency region but have difficulty with wider frequency regions, whereas our approach does not have such limitations. Furthermore, an important benefit that comes with our coherent wideband source localization approach is the ability to incorporate prior information on the frequency spectra of the sources. For example, in Fig. 15, where we performed incoherent processing, the spectra of the chirps have a jagged shape, due to the fact that we treat each frequency independently. To mitigate this artifact, in the coherent version of wideband processing, one could incorporate a prior on the continuity of the frequency spectra of the chirps. Another scenario where prior information on frequency could be particularly useful is for sources composed of multiple harmonics. In that case, a sparsity prior can be imposed on the frequency spectrum as well as on the spatial one. In Fig. 16,
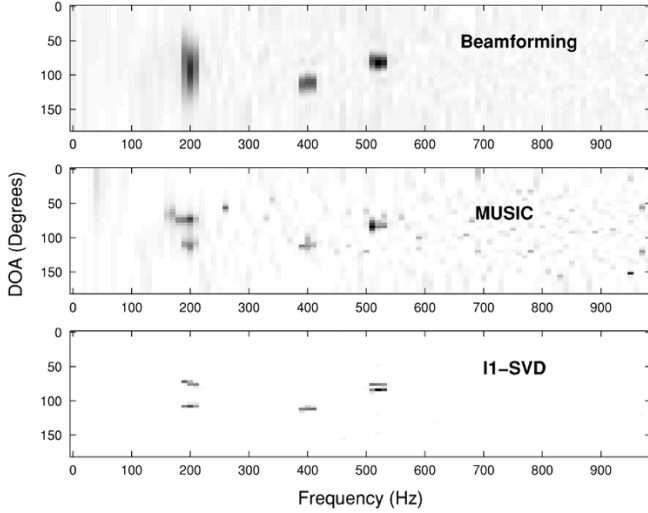
Fig. 16. Joint coherent processing of multiple harmonics with sparsity penalties on the spectra in the spatial and in frequency domain. Top: ℓ_1-SVD. Middle: Incoherent beamforming. Bottom: Incoherent MUSIC.

we look at three wideband signals consisting of one or two harmonics each. At DOA 76°, there are two harmonics with frequencies 200 and 520 Hz, at DOA 112°, there are again two harmonics with frequencies 200 and 400 Hz, and at DOA 84°, there is a single harmonic with frequency 520 Hz. Plot (a) shows results using conventional beamforming applied at each frequency (incoherently), plot (b) uses the MUSIC method applied at each frequency (incoherently), and plot (c) uses the coherent wideband version of ℓ_1-SVD. The results are displayed as intensity maps on a 2-D grid as a function of angle and frequency. Conventional beamforming merges the two well-separated peaks at 200 Hz, as well as the two closely spaced peaks at 520 Hz. MUSIC resolves the two peaks at frequency 200 Hz but merges the two at 520 Hz and shows some distortion due to noise. The coherent wideband version of ℓ_1-SVD resolves all five peaks and does not have any notable distortion due to noise.

IX. CONCLUSION

In this paper, we explored a formulation of the sensor array source localization problem in a sparse signal representation framework. We started with a scheme for source localization with a single snapshot and developed a tractable subspace-based ℓ_1-SVD method for multiple snapshots. The scheme can be applied to narrowband and to wideband scenarios. An efficient optimization procedure using SOC programming was proposed. We described how to mitigate the effects of the limitation of the estimates to a grid through an adaptive grid-refinement procedure and proposed an automatic method for choosing the regularization parameter using the constrained form of the discrepancy principle at high SNR. Finally, we examined various aspects of our approach, such as bias, variance, and the number of resolvable sources, using simulations. Several advantages over existing source localization methods were identified, including increased resolution, no need for accurate initialization, and improved robustness to noise, to a limited number of time samples and to correlation of the sources.

Some of the interesting questions for further research include an investigation of the applicability of greedy sparse signal representation methods, which have a lower computational cost, to source localization; a theoretical study of the bias and variance of our scheme; a detailed theoretical study of uniqueness and stability of sparse signal representation for the overcomplete bases that arise in source localization applications; a theoretical analysis of the multiple time-sample sparse signal representation problem; and applications of enforcing sparsity to spatially distributed or slowly time-varying sources.

APPENDIX
FORMULATING ℓ_1-SVD AS A SOC OPTIMIZATION PROBLEM

The general form of an SOC problem is

$$\min \mathbf{c}'\mathbf{x} \text{ such that } \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \text{and} \quad \mathbf{x} \in \mathbf{K}$$

where $\mathbf{K} = \mathbb{R}_+^N \times \mathbf{L}_1 \cdots \times \mathbf{L}_{N_L}$. Here, $\mathbb{R}_+^N$ is the N-dimensional positive orthant cone, and $\mathbf{L}_1, \dots, \mathbf{L}_{N_L}$ are SOCs.

First, to make our objective function in (13) linear, we use the auxiliary variables p and q and put the nonlinearity into the constraints by rewriting (13) as

$$\min p + \lambda q$$
$$\text{subject to } \|\mathbf{y}_{SV} - \mathbf{A}\mathbf{S}_{SV}\|_f^2 \le p, \quad \text{and} \quad \|\tilde{\mathbf{s}}^{(\ell_2)}\|_1 \le q \quad (17)$$

The vector $\tilde{\mathbf{s}}^{(\ell_2)}$ is composed of non-negative real values; hence, $\|\tilde{\mathbf{s}}^{(\ell_2)}\|_1 = \sum_{i=1}^N \tilde{s}_i^{(\ell_2)} = \mathbf{1}'\tilde{\mathbf{s}}^{(\ell_2)}$. The symbol $\mathbf{1}$ stands for an $N \times 1$ vector of ones. The constraint $\|\tilde{\mathbf{s}}^{(\ell_2)}\|_1 \le q$ can be rewritten as $\sqrt{\sum_{k=1}^K (s_i^{SV}(k))^2} \le r_i$, for $i = 1, \dots, N$, and $\mathbf{1}'\mathbf{r} \le q$, where $\mathbf{r} = [r_1, \dots, r_N]'$. In addition, let $\mathbf{z}_k = \mathbf{y}^{SV}(k) - \mathbf{A}\mathbf{s}^{SV}(k)$. Then, we have

$$\min p + \lambda q$$
$$\text{subject to } \|(\mathbf{z}_1', \dots, \mathbf{z}_K')\|_2^2 \le p, \quad \text{and} \quad \mathbf{1}'\mathbf{r} \le q$$
$$\text{where } \sqrt{\sum_{k=1}^K (s_i^{SV}(k))^2} \le r_i, \quad \text{for } i = 1, \dots, N. \quad (18)$$

The optimization problem in (18) is in the SOC programming form: We have a linear objective function and a set of quadratic, linear, and SOC constraints. Quadratic constraints can be readily represented in terms of SOC constraints; see [32] for details.

REFERENCES

[1] H. Krim and M. Viberg, "Two decades of array signal processing research. The parametric approach," in *IEEE Signal Process. Mag.*, vol. 13, Jul. 1996, pp. 67–94.

[2] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing—Concepts and Techniques*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[3] R. O. Schmidt, "A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1981.

[4] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[5] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–310, Feb. 1997.

[6] S. Sardy, P. Tseng, and A. Bruce, "Robust wavelet denoising," *IEEE Trans. Signal Process.*, vol. 49, no. 6, pp. 1146–1152, Jun. 2001.

[7] P. S. Bradley, O. L. Mangasarian, and W. N. Street, "Feature selection via mathematical programming," in *INFORMS J. Comput.*, vol. 10, 1998, pp. 209–217.

[8] M. Çetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 623–631, Apr. 2001.

[9] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. R. Statist. Soc.*, ser. B, vol. 58, pp. 267–288, Nov. 1996.

[10] M. D. Sacchi, T. J. Ulrych, and C. J. Walker, "Interpolation and extrapolation using a high-resolution discrete fourier transform," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 31–38, Jan. 1998.

[11] B. D. Jeffs, "Sparse inverse solution methods for signal and image processing applications," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, 1998, pp. 1885–1888.

[12] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[13] J. J. Fuchs, "Linear programming in spectral estimation. Application to array processing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 6, 1996, pp. 3161–3164.

[14] ——, "On the application of the global matched filter to DOA estimation with uniform circular arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 702–709, Apr. 2001.

[15] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 187–200, Jan. 1999.

[16] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," in *SIAM J. Sci. Comput.*, vol. 20, 1998, pp. 33–61.

[17] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[18] D. M. Malioutov, M. Cetin, and A. S. Willsky, "Optimal sparse representations in general overcomplete bases," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2004.

[19] A. J. Miller, *Subset Selection in Regression*. London, U.K.: Chapman and Hall, 2002.

[20] D. M. Malioutov. (2003, Jul.) A Sparse Signal Reconstruction Perspective for Source Localization With Sensor Arrays. M.S. thesis, Mass. Inst. Technol., Cambridge, MA. [Online]. Available: http://ssg.mit.edu/~dmm/publications/malioutov_MS_thesis.pdf

[21] D. L. Donoho and M. Elad, "Maximal sparsity representation via l1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, pp. 2197–2202, Mar. 2003.

[22] R. Gribonval and M. Nielsen, "Sparse representation in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[23] J. J. Fuchs, "More on sparse representations in arbitrary bases," in *Proc. 13th Symp. Syst. Ident.*, pp. 1357–1362.

[24] R. Gribonval and M. Nielsen, Highly sparse representations from dictionaries are unique and independent of the sparseness measure, 2003. submitted for publication.

[25] J. A. Tropp, "Greed is Good: Algorithmic Results for Sparse Approximation," Univ. Texas at Austin, ICES Rep. 03-04, Feb. 2003. submitted for publication.

[26] D. L. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presense of noise," *IEEE Trans. Inf. Theory*, submitted for publication.

[27] J. A. Tropp, Just relax: Convex programming methods for subset selection and sparse approximation, , Univ. Texas, Austin, ICES Rep. 04-04, Feb. 2004. submitted to *IEEE Trans. Inf. Theory*.

[28] D. L. Donoho, "Superresolution via sparsity constraints," *SIAM J. Math. Anal.*, vol. 23, pp. 1309–1331, 1992.

[29] J. J. Fuchs, "Detection and estimation of superimposed signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, 1998, pp. 1649–1652.

[30] B. D. Rao and K. Kreutz-Delgado, "Basis selection in the presense of noise," in *Proc. Thirty-Second Asilomar Conf. Signals, Syst., Comput.*, vol. 1, Nov. 1998, pp. 752–756.

[31] M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Its Applicat., Special Issue on Linear Algebra in Control, Signals, and Image Processing*, no. 284, pp. 193–228, 1998.

[32] A. Nemirovski and A. Ben Tal, *Lectures on Modern Convex Optimization. Analysis, Algorithms and Engineering Applications*. Philadelphia, PA: SIAM, 2001.

[33] J. S. Sturm. (2001) Using SeDuMi 1.02, a Matlab Toolbox for Optimization Over Symmetric Cones. Tilburg University, Dept. Econometrics, Tilburg, Netherlands. [Online]. Available: http://fewcal.kub.nl/~sturm

[34] V. A. Morozov, "On the solution of functional equations by the method of regularization," *Soviet Math. Dokl.*, vol. 7, pp. 414–417, 1966.

[35] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the l-curve," *SIAM Rev.*, vol. 34, pp. 561–580, Dec. 1992.

[36] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-Delgado, "Subset selection in noise based on diversity measure minimization," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 760–770, Mar. 2003.

[37] S. Sivanand, J. F. Yang, and M. Kaveh, "Time-domain coherent signal-subspace wideband direction-of-arrival estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, 1989, pp. 2772–2775.

Dmitry Malioutov (S'04) received the B.S. degree (summa cum laude) in electrical and computer engineering from Northeastern University, Boston, MA, in 2001 and the M.S. degree in 2003 from the Massachusetts Institute of Technology (MIT), Cambridge, where he is currently pursuing the Ph.D. degree.

His research interests include statistical signal and image processing, sensor array processing, sparse signal representation, convex optimization, and machine learning.

Müjdat Çetin (S'98–M'02) received the B.S. degree from Boügazi Çci University, Istanbul, Turkey, in 1993, the M.S. degree from the University of Salford, Manchester, U.K., in 1995, and the Ph.D. degree from Boston University, Boston, MA, in 2001, all in electrical engineering.

Since 2001, he has been a member of the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, where he now holds the title of Research Scientist. His research interests include statistical signal and image processing, inverse problems, biomedical signal processing, radar imaging, sensor array processing, and image segmentation.

Dr. Çetin has served in various organizational capacities, including Session Chair and Technical Program Committee Member for the SPIE Conference on Algorithms for Synthetic Aperture Radar Imagery, and Technical Program Committee member for a number of conferences, including the IEEE International Conference on Image Processing and the IEEE International Conference on Acoustics, Speech, and Signal Processing.

Alan S. Willsky (S'70–M'73–SM'82–F'86) joined the faculty of the Massachusetts Institute of Technology (MIT), Cambridge, in 1973 and is currently the Edwin Sibley Webster Professor of Electrical Engineering. He is a founder, member of the Board of Directors, and Chief Scientific Consultant of Alphatech, Inc. From 1998 to 2002, he served as a member of the U.S. Air Force Scientific Advisory Board. He has held visiting positions in England and France. He has delivered numerous keynote addresses and is co-author of the undergraduate text *Signals and Systems* (Englewood Cliffs, NJ: Prentice-Hall, 1996, Second ed.). His research interests are in the development and application of advanced methods of estimation and statistical signal and image processing. Methods he has developed have been successfully applied in a variety of applications including failure detection, surveillance systems, biomedical signal and image processing, and remote sensing.

Dr. Willsky has received several awards, including the 1975 American Automatic Control Council Donald P. Eckman Award, the 1979 ASCE Alfred Noble Prize, and the 1980 IEEE Browder J. Thompson Memorial Award. He has held various leadership positions in the IEEE Control Systems Society (which made him a Distinguished Member in 1988).

Feature-preserving regularization method for complex-valued inverse problems with application to coherent imaging

Müjdat Çetin
Sabanci University
Faculty of Engineering and Natural Sciences
34956, Istanbul, Turkey
E-mail: mcetin@sabanciuniv.edu

William C. Karl
Boston University
Multi-Dimensional Signal Processing Laboratory
Boston, Massachusetts 02215

Alan S. Willsky, MEMBER SPIE
Massachusetts Institute of Technology
Laboratory for Information and Decision Systems
Cambridge, Massachusetts 02139

Abstract. We propose a method for feature-preserving regularized reconstruction in coherent imaging systems. In our framework, image formation from measured data is achieved through the minimization of a cost functional, designed to suppress noise artifacts while preserving features such as object boundaries in the reconstruction. The cost functional includes nonquadratic regularizing constraints. Our formulation effectively deals with the complex-valued and potentially random-phase nature of the scattered field, which is inherent in many coherent systems. We solve the challenging optimization problems posed in our framework by developing and using an extension of half-quadratic regularization methods. We present experimental results from three coherent imaging applications: digital holography, synthetic aperture radar, and ultrasound imaging. The proposed technique produces images where coherent speckle artifacts are effectively suppressed, and important features of the underlying scenes are preserved. © *2006 Society of Photo-Optical Instrumentation Engineers.* [DOI: 10.1117/1.2150368]

1 Introduction

This paper addresses image reconstruction problems in coherent imaging. Coherent imaging is based on recording spatial and/or temporal variations in both the intensity of a scattered field and its phase.[1] Many microwave, optical, and acoustic sensing applications use coherent imaging, and particular modalities include synthetic-aperture radar (SAR), holography, sonar, ultrasound, and laser imaging, among others. In both coherent and incoherent imaging tasks, reconstruction of an image from observed data is often an ill-posed inverse problem. Solution of such inverse problems can be achieved through regularization methods, which turn the problem into a well-posed one and prevent the amplification of measurement noise during the reconstruction process. However, one limitation of straightforward regularization methods, such as Tikhonov regularization,[2] is the suppression of important features in the resulting imagery, such as edges. Recently this issue has been successfully addressed by feature-preserving regularization techniques in incoherent imaging applications, such as restoration of blurred and noisy optical images[3] and reconstruction in x-ray tomography.[4]

Coherent image reconstruction poses additional challenges that do not appear in incoherent imaging. First, the signals involved are in general complex-valued. Furthermore, in many problems, including SAR and holography of diffuse objects, the phase of the scattered field is a highly random quantity.* This leads to two complications. First, due to constructive and destructive interference of scatterers within a resolution cell, conventional coherent images suffer from speckle artifacts. (Speckle appears when the surface being imaged has roughness at the scale of the illuminating wavelength.) Second, due to the complex-valued and possibly random-phase nature of the fields, straightforward application of image reconstruction methods originally designed for incoherent imaging may not produce accurate reconstructions, as we experimentally demonstrate in Sec. 3.

To address these challenges, we propose a feature-preserving regularization method specifically for coherent imaging tasks. The approach we present involves the minimization of a cost functional that contains *nonquadratic regularization* constraints. Such nonquadratic constraints have been shown to lead to feature preservation by preferring reconstructions that are *sparse* in terms of the features of interest.[6] Our framework is general enough to handle various features (as we demonstrate later), but for the sake of concreteness at this point, let us assume that the features of interest are the boundaries between distinct physically meaningful regions in the scene. The goal then is to reconstruct images where various imaging artifacts and noise are suppressed, while object and region boundaries (edges) are

---

*This property is known to enable high-quality reconstructions from limited Fourier-offset data in coherent imaging.[5] For this reason, Fourier transform holograms are often constructed using a diffuser to impart essentially random phase to each point in the original scene before recording.

preserved. The regularization constraints in our framework achieve artifact suppression by imposing smoothness on the *magnitudes* of the reconstructed complex-valued field reflectivities (or transmission coefficients). The nonquadratic aspect of these regularizing functionals leads to edge preservation, similar to the case in incoherent imaging problems.[3,4] To solve the resulting optimization problems, we provide a formal extension of half-quadratic regularization techniques[7] to complex-valued, random-phase fields. This constitutes the major technical contribution of our work.

There are a number of publications that are related to some of the coherent-imaging issues that we address. The implications of the random-phase nature of coherent images in terms of the quality of the reconstructions has been analyzed in Refs. 5 and 8. The work in Ref. 9 presents a maximum likelihood technique for reconstructing complex-valued, random-phase images from Fourier-offset data using the expectation-maximization algorithm. Bayesian techniques have been used for filtering complex-valued, speckled images in Ref. 10, and for ultrasound Doppler spectral analysis based on autoregressive models in Ref. 11. A technique for image reconstruction from noisy digital holograms based on the method of projection onto convex sets (POCS) has been developed in Ref. 12. These last three papers are somewhat related to our approach in that they use regularizing constraints. A number of more recent publications have a closer relation to our perspective for coherent imaging, in particular in their emphasis on preservation of edges or other features. A Bayesian approach for the nonlinear inverse scattering problem of tomographic imaging using microwave or ultrasound probing has been proposed in Ref. 13. In Refs. 14 and 15, maximum-entropy regularization has been used for image reconstruction from sparsely sampled coherent field data. The work in Ref. 16 proposes a regularized autoregressive model for spectral estimation, with application to medical ultrasonic radio-frequency images. Another method for spectral estimation involves regularization through a circular Gibbs-Markov model.[17] A statistical deconvolution technique for diffuse ultrasound scattering has been proposed in Ref. 18, where sampling techniques are used for inference. In Ref. 19, anisotropic diffusion[20] has been used for ultrasound speckle reduction and coherence enhancement. The total variation-based regularization method proposed in Ref. 21 has been applied to coherent imaging, in particular to near-field acoustic holography. Finally, in Ref. 22, a penalized-likelihood image reconstruction technique has been proposed for image-plane holography, which uses incoherent illumination.

Our approach is significantly different from this body of previous work in a number of ways. First, we consider the random-phase aspect (and deal with the effects of speckle) much more explicitly than any of the previous papers on inverse problems in coherent imaging. Second, the structures of the energy functionals used in our framework are quite different from what has been used in previous work, and this structure allows the use of a variety of regularizing constraints within a single framework. Third, the algorithm we use for optimization, namely an extension of half-quadratic regularization, is new. We demonstrate the performance of the proposed method on examples from a number

of coherent imaging applications. With enhanced speckle and artifact suppression, as well as feature preservation, the images produced by our method appear to yield more accurate reconstructions than conventional coherent imaging techniques.

In Sec. 2, we present our nonquadratic regularization-based approach. We first develop the method with ℓ_p-norm-based potential functionals, and then extend it to other nonquadratic potentials. Section 3 contains the experimental results, and we conclude in Sec. 4.

2 Nonquadratic Regularization for Complex-Valued Problems

This section contains the description of the nonquadratic technique we propose in this paper. We start by describing the general form of the observation models we consider. We then formulate an optimization problem for coherent imaging, which involves a cost functional based on ℓ_p norms. To minimize this cost functional, we propose an algorithm based on half-quadratic regularization, and provide a statistical interpretation of this strategy. Finally we generalize our method to incorporate nonquadratic cost functionals other than ℓ_p norms.

2.1 Observation Model

In this paper, we are concerned with inverse problems in which the sensor measurements y are related to the underlying, unknown field f, through a Fredholm integral equation of the first kind:

$$y(\mathbf{x}) = \int_{\Omega} T(\mathbf{x}, \mathbf{x}')f(\mathbf{x}')\,\mathrm{d}\mathbf{x}' + w(\mathbf{x}), \tag{1}$$

where Ω is the spatial region of interest for the reconstruction, and w is additive measurement noise. The argument of f corresponds to 2-D or 3-D spatial coordinates, and the arguments of y and w depend on the domain of the measurements in specific applications.

We assume that the integral kernel T, which models the relationship between the underlying field and the measured data, is known. For example, T may be a band-limited, possibly frequency-offset Fourier transform operator, where the physics of the problem, the sensor parameters, and the observation geometry determine the exact structure. Another example for T, used in tomographic imaging modalities, is projection-type operators, related to the Radon transform.[23] Yet another form arising in many applications is convolutional operators. For some particular observation models that are of interest in our work (and that we use in our experimental analysis), see Refs. 24 for digital holography, Refs. 25 and 26 for SAR, and Refs. 14, 27, and 28 for ultrasound.

In many coherent imaging applications, which involve, e.g., multiple scattering and other second-order phenomena, the exact equations governing the observation process are actually nonlinear. In such scenarios, approximate linear observation models as in Eq. (1) can be obtained through first-order solutions, which exclude all but primary scattering. Such linear models include the well-known Born approximation and the physical optics approximation. These linear approximations yield acceptable results in many
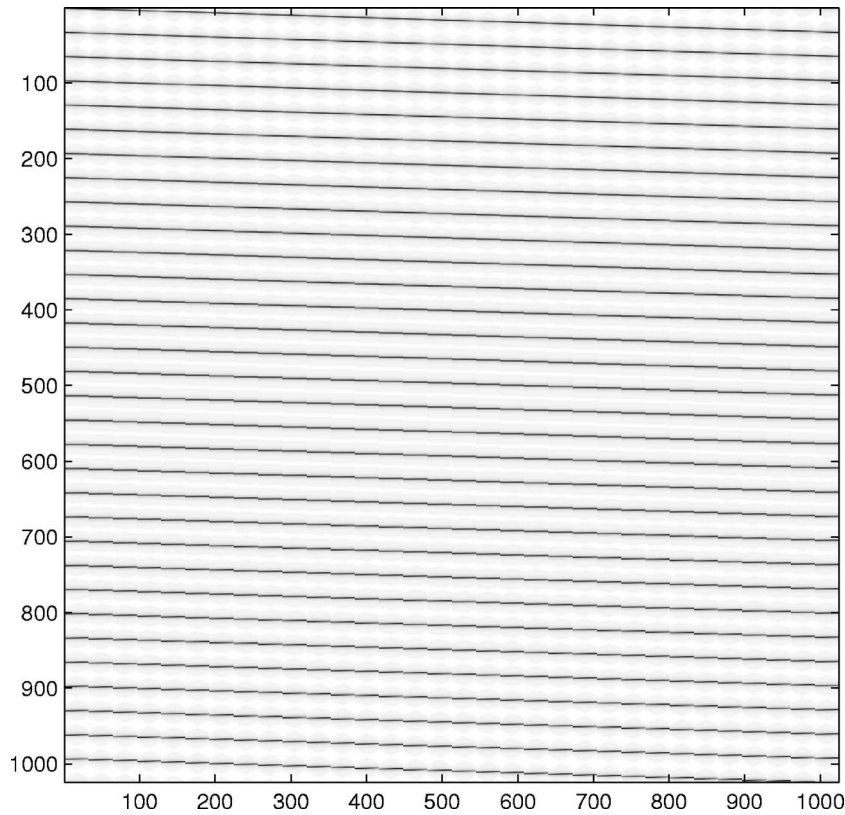
Fig. 1 Grayscale plot (black corresponds to the maximum value, and white to the minimum) of the magnitude of the elements in a SAR projection matrix for a 32×32 field. The radar in this example operates in the X band with a center frequency of 10 GHz, and the underlying scene is viewed through an angular span of 2.3 deg.

practical situations, where our techniques are directly applicable. On the other hand, it is certainly of interest to develop inversion techniques based on more accurate nonlinear models. While we address only linear problems in this paper, the key ideas we present are potentially useful in nonlinear problems as well, and our method could be generalized to such cases.

In practice, we discretize the relationship in Eq. (1) and use the following model for the coherent observation process:

$$\mathbf{y} = \mathbf{T}\mathbf{f} + \mathbf{w}, \tag{2}$$

where $\mathbf{y}, \mathbf{f}$, and $\mathbf{w}$ are the sampled data, the unknown image, and noise, respectively, all column-stacked as vectors. Similarly, $\mathbf{T}$ is a matrix representing the discrete observation kernel. To provide some flavor of such discrete operators, in Fig. 1 we illustrate a tomographic projection operator that arises in one of the applications of interest in this paper, namely, SAR. The operator is complex-valued, and we show only the magnitudes of the elements of the matrix as a grayscale plot. Each column of the matrix corresponds to one spatial location in the underlying image, and describes how the reflectivity at that location contributes to the projectional radar observations. Each row of the matrix corresponds to one particular data point (one sample in the discretized radar return at a particular observation angle), and describes the effect of various spatial locations in the scene on that data point.

Given the observation model in Eq. (2), the objective is to obtain a reconstruction of $\mathbf{f}$, based on the data $\mathbf{y}$. Conventional image formation techniques vary depending on the particular modality and sensor model, and include algorithms based on beamforming, filtered backprojection, and inverse Fourier transformation, among others.

2.2 Cost Functional Based on ℓ_p Norms

We propose to find the reconstructed image $\hat{\mathbf{f}}$ as the minimizer of the following cost functional:

$$J_0(\mathbf{f}) = \|\mathbf{y} - \mathbf{T}\mathbf{f}\|_2^2 + \lambda \||\mathbf{D}|\mathbf{f}|\|_p^p, \tag{3}$$

where $\|\cdot\|_p$ denotes the ℓ_p norm, $\mathbf{D}$ is a matrix to be described below, $|\mathbf{f}|$ denotes the vector of magnitudes of the complex-valued vector $\mathbf{f}$, and $\lambda, p < 2$ are scalar parameters.[†] Note that the formulation of Eq. (3) takes into account the forward model $\mathbf{T}$ and starts from the observed sensor data $\mathbf{y}$, and hence is not simply a postprocessing of a formed image.

The first term of $J_0(\mathbf{f})$ in Eq. (3) is a data fidelity term, while the second term incorporates prior information regarding both the behavior of the field $\mathbf{f}$ and the nature of the features of interest in the resulting reconstructions. In par-

---

[†]When $p < 1$, the triangle inequality is not satisfied and it would be more precise to use the term "quasi-norm" rather than "norm." However, we ignore this subtlety and use the term "ℓ_p norm" for any value of p.

ticular, the nonquadratic structure of the second term provides feature preservation,[3,4] where the matrix $\mathbf{D}$ determines the kind of features to be preserved. For example, if we are interested in reconstructing images consisting of spatially extended objects and regions, with slowly varying physical properties (such as reflectivities) within the regions, then a good choice for $\mathbf{D}$ is a discrete approximation to the 2-D spatial derivative operator (gradient). With this choice, the second term in Eq. (3) becomes a piecewise smoothness constraint, imposing smoothness within regions and allowing sharp transitions across the region boundaries, leading to edge preservation. In Sec. 3, we show examples demonstrating the use of this choice of $\mathbf{D}$ on digital holography and SAR imaging. For a discussion of the structure of 2-D discrete derivative operators, see the Appendix (Sec. 5.1).

While edge-preserving reconstruction is of interest in many coherent-imaging tasks, one might also be interested in other features. For example, rather than spatially extended objects, an application might involve imaging spatially localized scatterers. In that case, we would be interested in preserving the scattering amplitudes of the strong scatterers in the scene, while suppressing noise and artifacts. In our framework, this could be achieved by choosing $\mathbf{D}$ to be an identity operator in Eq. (3). Such constraints have been shown to lead to superresolution.[29] In Sec. 3, we show examples demonstrating the use of this choice of $\mathbf{D}$ on ultrasound imaging.

In order to avoid problems due to nondifferentiability of the ℓ_p norm around the origin when $p \leq 1$, we use a smooth approximation to the ℓ_p-norm in Eq. (3).[3] This leads to the following slightly modified cost functional to be used in practice for numerical purposes:

$$J(\mathbf{f}) = \|\mathbf{y} - \mathbf{Tf}\|_2^2 + \lambda \sum_{i=1}^{M} \left[|(\mathbf{D}|\mathbf{f}|)_i|^2 + \epsilon \right]^{p/2}, \tag{4}$$

where $\epsilon \geq 0$ is a small constant, $(\cdot)_i$ denotes the i'th element of a vector, and M is the length of the vector $\mathbf{D}|\mathbf{f}|$. Note that $J(\mathbf{f}) \rightarrow J_0(\mathbf{f})$ as $\epsilon \rightarrow 0$.[‡]

Nonquadratic regularizing constraints such as ℓ_p norms have previously been shown to produce feature-preserving solutions in problems such as image restoration[3] and x-ray tomography,[4] where the signals involved are real-valued. In contrast, we are interested in coherent systems such as SAR and holography, where the processed signals are complex-valued. In many cases of interest, the phase of the unknown complex-valued field $\mathbf{f}$ is highly random, and uncorrelated with the phase at neighboring pixels. Based on this observation, in such coherent imaging problems, regularizing constraints such as smoothness should be applied explicitly to the *magnitudes* $|\mathbf{f}|$ of the complex-valued reflectivities $\mathbf{f}$. In our framework, this is achieved through the expression $\mathbf{D}|\mathbf{f}|$ in Eq. (4). This nonlinearity in $\mathbf{f}$ makes the optimization problem more challenging than those arising in incoherent imaging applications. In the next subsection, we pro-

pose an extension of half-quadratic regularization methods[7] to complex-valued, random-phase fields for achieving efficient and robust numerical solution of the optimization problems of the form (4), posed in our framework.

2.3 Half-Quadratic Regularization for Coherent Imaging

The main idea in half-quadratic regularization is to introduce and optimize a new cost functional, which has the same minimum as the original nonquadratic cost functional [in our case, $J(\mathbf{f})$], but one which can be manipulated with linear algebraic methods. In incoherent imaging applications, such a new cost functional is obtained by augmenting the original cost functional with an *auxiliary* vector.

Currently available half-quadratic regularization methods designed for incoherent imaging cannot handle the more complicated structure of the optimization problems involved in coherent imaging. In order to deal with such complications, we propose using two auxiliary vectors, $\mathbf{b}$ and $\mathbf{s}$, matched to the structure of the problem, to form an augmented cost functional $K(\mathbf{f}, \mathbf{b}, \mathbf{s})$ which satisfies

$$\inf_{\mathbf{b}, \mathbf{s}} K(\mathbf{f}, \mathbf{b}, \mathbf{s}) = J(\mathbf{f}). \tag{5}$$

In particular, we construct $K(\mathbf{f}, \mathbf{b}, \mathbf{s})$ in such a way that it is quadratic in $\mathbf{f}$ (hence the name half-quadratic) and easy to minimize in $\mathbf{b}$ and $\mathbf{s}$. Then the minimization of $K(\mathbf{f}, \mathbf{b}, \mathbf{s})$ can be performed through a block coordinate descent approach.

Now, let us consider our particular cost functional $J(\mathbf{f})$ of Eq. (4). We can show that the following augmented cost functional $K(\mathbf{f}, \mathbf{b}, \mathbf{s})$ satisfies the relationship (5) for the particular $J(\mathbf{f})$ of Eq. (4) (see Appendix, Sec. 5.2):

$$K(\mathbf{f}, \mathbf{b}, \mathbf{s}) = \|\mathbf{y} - \mathbf{Tf}\|_2^2 + \lambda \sum_{i=1}^{M} \left\{ \mathbf{b}_i \left[|(\mathbf{DSf})_i|^2 + \epsilon \right] \right.$$
$$\left. + \left(\frac{p}{2\mathbf{b}_i} \right)^{\frac{p}{2-p}} \left(1 - \frac{p}{2} \right) \right\}, \tag{6}$$

where

$$\mathbf{S} = \mathrm{diag}\{\exp(-j\mathbf{s}_l)\}, \tag{7}$$

with $\mathbf{s}_l$ the l'th element of the vector $\mathbf{s}$, and $\mathrm{diag}\{\cdot\}$ a diagonal matrix whose l'th diagonal element is given by the expression inside the braces. Due to Eq. (5), $J(\mathbf{f})$ and $K(\mathbf{f}, \mathbf{b}, \mathbf{s})$ share the same minima in $\mathbf{f}$. Note that $K(\mathbf{f}, \mathbf{b}, \mathbf{s})$ is a quadratic function with respect to $\mathbf{f}$.[§] We benefit from the half-quadratic structure through the use of an iterative block coordinate descent method on $K(\mathbf{f}, \mathbf{b}, \mathbf{s})$, in order to find the field $\hat{\mathbf{f}}$ that also minimizes $J(\mathbf{f})$:

$$\hat{\mathbf{s}}^{(n+1)} = \arg \min_{\mathbf{s}} K(\hat{\mathbf{f}}^{(n)}, \hat{\mathbf{b}}^{(n)}, \mathbf{s}), \tag{8}$$

[‡]Note that there is still some nondifferentiability left in $J(\mathbf{f})$, due to $|\mathbf{f}|$. One could in principle apply a similar smooth approximation for this term. However, to keep the notation simple, we ignore this subtlety in our development. One could avoid any practical difficulties this might cause simply by defining the phase at the origin of the complex plane to be zero.

[§]We have obviously omitted the recipe for finding a valid $K(\mathbf{f}, \mathbf{b}, \mathbf{s})$ from $J(\mathbf{f})$ here. We just want to point out that, given any feature-preserving cost functional $J(\mathbf{f})$, the augmented cost functional can be found by using convex duality relationships, and we refer the interested reader to Ref. 7.

$$\hat{\mathbf{b}}^{(n+1)} = \arg \min_{\mathbf{b}} K(\hat{\mathbf{f}}^{(n)}, \mathbf{b}, \hat{\mathbf{s}}^{(n+1)}), \qquad (9)$$

$$\hat{\mathbf{f}}^{(n+1)} = \arg \min_{\mathbf{f}} K(\mathbf{f}, \hat{\mathbf{b}}^{(n+1)}, \hat{\mathbf{s}}^{(n+1)}), \qquad (10)$$

where n denotes the iteration number. Using results from Sec. 5.2, we obtain

$$\hat{\mathbf{s}}_i^{(n+1)} = \phi[(\hat{\mathbf{f}}^{(n)})_i], \qquad (11)$$

$$\hat{\mathbf{b}}_i^{(n+1)} = \frac{p}{2[(\mathbf{D}\hat{\mathbf{S}}^{(n+1)}\hat{\mathbf{f}}^{(n)})_i^2 + \epsilon]^{1-p/2}}, \qquad (12)$$

$$[\mathbf{T}^H\mathbf{T} + \lambda(\hat{\mathbf{S}}^{(n+1)})^H\mathbf{D}^T\mathrm{diag}\{\hat{\mathbf{b}}_i^{(n+1)}\}\mathbf{D}\hat{\mathbf{S}}^{(n+1)}]\hat{\mathbf{f}}^{(n+1)} = \mathbf{T}^H\mathbf{y}, \quad (13)$$

where $\phi[z]$ denotes the phase of the complex number z. We can substitute Eqs. (11) and (12) into (13) to obtain a single iterative expression for $\hat{\mathbf{f}}^{(n+1)}$, which would then constitute the overall iterative algorithm.

Note that each iteration in Eq. (13) requires the solution of a set of linear equations for the unknown $\hat{\mathbf{f}}^{(n+1)}$. The coefficient matrix of this set of equations is Hermitian, positive semidefinite, and usually sparse. Hence these equations may themselves be efficiently solved using iterative approaches. We use the conjugate gradient (CG) algorithm for this solution, and terminate it when the ℓ_2 norm of the relative residual becomes smaller than a threshold $\delta_{\mathrm{CG}} > 0$.[30] We run the iteration (13) until $\|\hat{\mathbf{f}}^{(n+1)} - \hat{\mathbf{f}}^{(n)}\|_2^2 / \|\hat{\mathbf{f}}^{(n)}\|_2^2 < \delta$, where $\delta > 0$ is a small constant. In the Appendix (Sec. 5.3), we show that this algorithm is convergent in terms of the cost functional. For algorithms of this type, stronger results on the convergence of the iterates exist,[4,31] requiring certain assumptions on the nature of the cost functionals[4] or on the nature of the local minima.[31] For the specific algorithm we present here, we have not yet carried out such a more detailed analysis. In our algorithm, we use a stopping criterion based on the relative change in the iterates $\hat{\mathbf{f}}^{(n)}$ as stated, and we have not run into any convergence problems in practice. In general, the algorithm appears to be reaching a local minimum from any initialization.

2.4 Statistical Interpretation of Half-Quadratic Regularization

It is well known that optimization problems of the form in Eq. (3) can also be interpreted as statistical estimation problems (see, e.g., Ref. 32). In particular, the same optimization problem is reached when we try to find the maximum *a posteriori* (MAP) estimate of the field $\mathbf{f}$ based on the data $\mathbf{y}$ using a Gaussian, independent identically distributed noise model, together with a generalized Gaussian prior model for the field reflectivity magnitudes, where the spatial dependence structure is governed by the matrix $\mathbf{D}$. The phase distribution is assumed to be uniform and spatially independent. As an example, when $p = 1$, we have a Laplacian prior model for the field magnitudes. This heavy-tailed nature of the prior distribution is what leads to preservation

Table 1 Families of potential functionals used. Here p is a parameter determining the shape of the functionals, and ϵ is a small smoothing constant.

$\psi_1(x)$	$(x^2 + \epsilon)^{p/2}$
$\psi_2(x)$	$\dfrac{(x^2 + \epsilon)^{p/2}}{1 + (x^2 + \epsilon)^{p/2}}$
$\psi_3(x)$	$\log[1 + (x^2 + \epsilon)^{p/2}]$

of features such as edges. Note that the prior distribution here is *non-Gaussian*, and spatially *stationary*.

Now, let us interpret our half-quadratic regularization-based algorithm statistically. First note that the cost functional in Eq. (6) is a quadratic function of the field $\mathbf{f}$. Consequently, the coordinate-descent-based minimization in Eqs. (11)–(13) essentially solves a sequence of quadratic minimization problems for the field [although this is not explicitly shown, it might be observed from the linear structure of the iteration in Eq. (13)]. However, the quadratic problems contain field-dependent weights involving the auxiliary vectors $\mathbf{b}$ and $\mathbf{s}$. From an estimation standpoint, we essentially have a Gaussian prior for the field, but the distribution is nonstationary due to the field-dependent weighting, which is adaptively determined. Hence, the half-quadratic regularization-based algorithm might be viewed as replacing the original stationary, non-Gaussian problem with a series of nonstationary but Gaussian problems.

2.5 Extension to Other Nonquadratic Functionals

In Sec. 2.2, we have formulated the image reconstruction problem using a particular family of regularizing functionals, namely ℓ_p norms. We now generalize our framework and iterative algorithm to incorporate a wider range of potentially useful choices, which have previously found use in incoherent image restoration and reconstruction problems.[4,7,33,34] To this end, let us consider the following general form for the cost functional:

$$J(\mathbf{f}) = \|\mathbf{y} - \mathbf{T}\mathbf{f}\|_2^2 + \lambda \sum_i \psi((\mathbf{D}|\mathbf{f}|)_i), \qquad (14)$$

where ψ denotes the regularizing functional.

Three particular classes of functionals ψ we consider in this paper are shown in Table 1.[‖] Note that the use of ψ_1 leads to constraints in terms of approximate ℓ_p norms, which is precisely what we have discussed in Sec. 2.2. The potential functional ψ_2 is based on previous work in Ref. 33. Special cases of ψ_2 for $p = 1$ and $p = 2$ yield the potential functionals used in Refs. 7 and 4, respectively. Finally, ψ_3 is a generalized version of the potential functional proposed in Ref. 34. Note that these potential functionals can more generally be expressed in terms of x/Δ, where Δ is a scaling parameter. We use a fixed Δ, and omit it in our analysis for notational simplicity.

[‖]One might subtract an appropriate constant from each potential functional to set $\psi_k(0) = 0$ ($k = 1, 2, 3$); however, we have chosen not to do so in Table 1, to keep the notation simpler.

**Table 2** The updates for the auxiliary variable **b** for each of the three potential functionals

Potential functional	Associated $\hat{\mathbf{b}}_i^{(n+1)}$
ψ_1	$\dfrac{p/2}{[(\mathbf{D}\hat{\mathbf{S}}^{(n+1)}\hat{\mathbf{f}}^{(n)})_i^2 + \epsilon]^{1-p/2}}$
ψ_2	$\dfrac{p/2}{[(\mathbf{D}\hat{\mathbf{S}}^{(n+1)}\hat{\mathbf{f}}^{(n)})_i^2 + \epsilon]^{1-p/2}\{[(\mathbf{D}\hat{\mathbf{S}}^{(n+1)}\hat{\mathbf{f}}^{(n)})_i^2 + \epsilon]^{p/2} + 1\}^2}$
ψ_3	$\dfrac{p/2}{[(\mathbf{D}\hat{\mathbf{S}}^{(n+1)}\hat{\mathbf{f}}^{(n)})_i^2 + \epsilon]^{1-p/2}\{[(\mathbf{D}\hat{\mathbf{S}}^{(n+1)}\hat{\mathbf{f}}^{(n)})_i^2 + \epsilon]^{p/2} + 1\}}$

We minimize $J(\mathbf{f})$ in Eq. (14) by using the half-quadratic regularization-based coordinate descent strategy in Eqs. (8)–(10). This requires finding and using the augmented cost functional $K(\mathbf{f},\mathbf{b},\mathbf{s})$ that satisfies the condition in Eq. (5) for the particular potential functional ψ used in $J(\mathbf{f})$ of Eq. (14). For the sake of brevity, we do not give the expressions for $K(\mathbf{f},\mathbf{b},\mathbf{s})$ for each of the potential functionals in Table 1, but rather mention how the iterations for the ℓ_p-norm case, given by Eqs. (11)–(13), would be affected by the use of a different functional. In fact, the only modification needed in the iterative algorithm of Eqs. (11)–(13) is the update for $\hat{\mathbf{b}}_i^{(n+1)}$ in Eq. (12). Table 2 shows the form of these updates for the three potential functionals of Table 1. Note that the framework we have presented is not limited to the three specific potential functionals we have used as examples, and other functionals might be used as well.

3 Experimental Results

We demonstrate the performance of our techniques on three imaging applications: digital holography, SAR, and ultrasound. For particular sensor models in these applications, see Refs. 14, 24, and 25. In the cost functional of Eq. (4), we find that values of p around 1 appear to yield good results for the applications we consider here. As a result, we use $p=1$ in all of our experimental results in this paper. We choose the hyperparameter λ, which appears in the cost functional $J(\mathbf{f})$ of Eq. (4), based on subjective qualitative assessment of the formed imagery. We set the approximation parameter in the nonquadratic potentials in Table 1 to be $\epsilon=10^{-5}$, which is small enough not to affect the behavior of the solution. For the termination condition of our iterative algorithm, we use $\delta=10^{-6}$ and a CG tolerance of $\delta_{CG} = 10^{-3}$.

Figure 2 contains the results of currently available methods for a holography experiment. The magnitude of the underlying complex-valued scene is shown in Fig. 2(a). The phase of the scene at each pixel is uniformly distributed, and uncorrelated with the phase at other pixels. We consider the case of Fraunhofer diffraction[1,24] and compute a band-limited Fourier hologram, which constitutes the measured data. The amount of data we have after bandlimitation is equal to 76% of the hologram data that would be needed to form a full-resolution reconstruction of the original image. The image in Fig. 2(b) is the magnitude of the conventional reconstruction from the hologram. This result is dominated by coherent speckle artifacts. We now

(a)   (b)
(c)   (d)
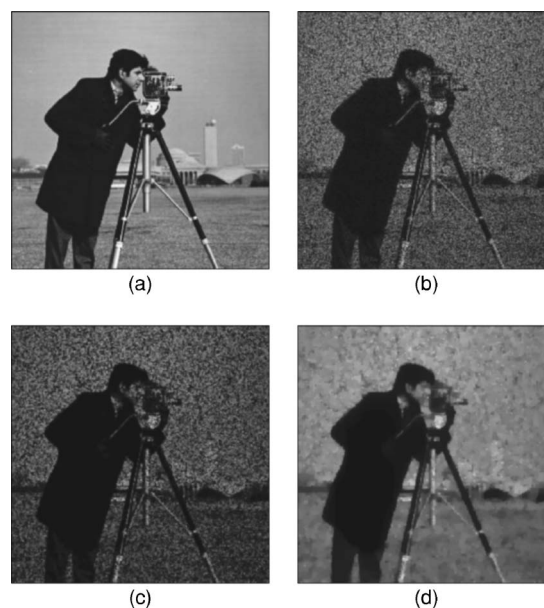
Fig. 2 Reconstruction of an image from its band-limited Fourier hologram using currently available techniques. (a) Original scene. (b) Conventional reconstruction. (c) Reconstruction by an edge-preserving regularization method designed for incoherent imaging. (d) Postprocessing of the conventionally reconstructed image by an anisotropic diffusion-based method.

show how incoherent image-processing techniques can fail in this problem. In Fig. 2(c), we show the result of an incoherent edge-preserving reconstruction method. In particular we use nonquadratic regularization with ℓ_p-norm-based constraints.[3,4] Since such techniques have been designed for real-valued signals, they are not able to treat the magnitude and phase components properly. This leads to some smoothing in the real and imaginary components of the field; however, a speckle-dominated magnitude image is produced, which shows only minor improvement over the conventional image of Fig. 2(b). In Fig. 2(d), we present the result of applying a variant of anisotropic diffusion[20] to the magnitude of the conventionally reconstructed image. Some speckle suppression seems to have been achieved; however, a significant amount of detail in the scene has been lost.

In Fig. 3, we present the results of the technique we have proposed in Sec. 2, with each of the three regularizing potentials from Table 1, and $p=1$. In this experiment, we choose **D** to be a discrete approximation to the 2-D spatial derivative operator. With suppressed speckle and preserved edges, our method provides what appears to be an accurate reconstruction of the original scene in Fig. 2(a). These results demonstrate the power of our model-based coherent image reconstruction approach as compared to standard coherent image formation [Fig. 2(b)], incoherent edge-preserving regularization [Fig. 2(c)], and anisotropic diffusion-based postprocessing for image enhancement [Fig. 2(d)].

For the remaining examples, we only present images produced by conventional imaging and our nonquadratic regularization-based method. An additional analysis similar
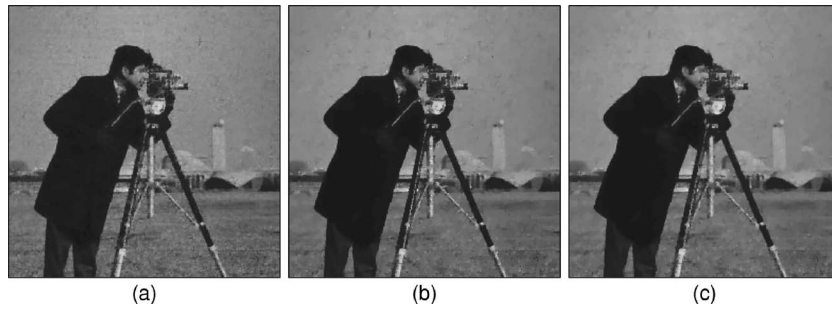
Fig. 3 Reconstruction of an image from its band-limited Fourier hologram by the technique proposed in Sec. 2, with the following choices of regularizing functionals from Table 1, and $p=1$: (a) ψ_1, (b) ψ_2, (c) ψ_3.

to that carried out for the digital holography example of Fig. 2 and 3 yields qualitatively very similar results.

Our next example is from X-band SAR imaging, where we use a tomographic observation model.[25] Figure 4(a) contains a conventional SAR image of three vehicles in a field containing some trees. Speckle artifacts, clearly visible in this reconstruction, make, e.g., automatic segmentation of SAR images very challenging. In contrast, the images produced by our method (with $p=1$ and **D** being a derivative operator), shown in Fig. 4(b)–4(d) for different regularizing potentials ψ, produce regions (vehicle, tree, shadow, background) that appear to be more easily separable.

Our final results are from ultrasound imaging motivated by the application of nondestructive evaluation (NDE). One of the goals in nondestructive evaluation is to image the internal structure of homogeneous materials to detect material defects such as cracks. We present experimental results based on data collected at the Large Ultrasound Test

Fig. 4 (a) Conventional SAR image of a scene. (b),(c),(d) Reconstructions produced by the technique proposed in Sec. 2, with the following choices of regularizing functionals from Table 1, and $p=1$: (b) ψ_1, (c) ψ_2, (d) ψ_3.

Facility[35] at Boston University. The goal in this experimental setup is to image the cross section of an aluminum object (modeling the crack) immersed in a tank full of water (modeling the homogeneous material). Data are collected in a monostatic data acquisition configuration by mechanically scanning a single transducer through a set of aperture coordinates above the tank. At each data collection point, we record a broadband echo signal. For the experiments reported here, we only use frequency-domain data at a temporal frequency of 730 kHz, although our approach could also use data at multiple frequencies. For the mathematical model relating the underlying image to the observed data, we use the physical optics approximation, as in Ref. 14. This leads to a Green's function, or a complex-valued point spread function (PSF), which we use to construct the matrix **T** in Eq. (2). This theoretical observation model appears to be in good agreement with the experimental PSF we have obtained using a spherical point target in our experimental setup. Further details of this experimental setup are beyond the scope of the current paper, and will be described elsewhere. Let us now start presenting our image reconstruction results. The synthetic image in Fig. 5(a) shows the U-shaped cross section of the aluminum object, based on the true dimensions of the object, and its actual relative location within the viewing geometry. This synthetic image is just to help visualize the "underlying true field" in this experiment, and the results we present next are based on measured data and not on synthetically generated data. In Fig. 5(b), we show a conventional image, reconstructed using a regularized pseudoinverse technique.[36] Such techniques are widely used in a variety of inverse problems. This image exhibits some artifacts, making it difficult to determine the shape of the imaged object (hence the shape and structure of the crack in NDE). In this application, the goal is to image narrow cracks rather than spatially distributed objects; hence in our methods we use **D=I** in Eqs. (4) and (14). Our technique ($p=1$) produces the images in Fig. 5(c)–5(e) where artifacts are reduced, and the shape of the aluminum object is preserved.

4 Conclusions

We have presented an optimization-based method for image formation in coherent systems. Our approach is based on cost functionals that are extensions of nonquadratic regularization techniques. The cost functionals are constructed
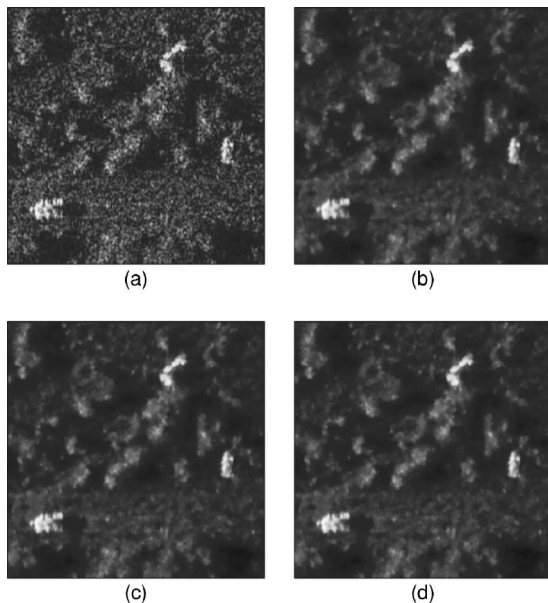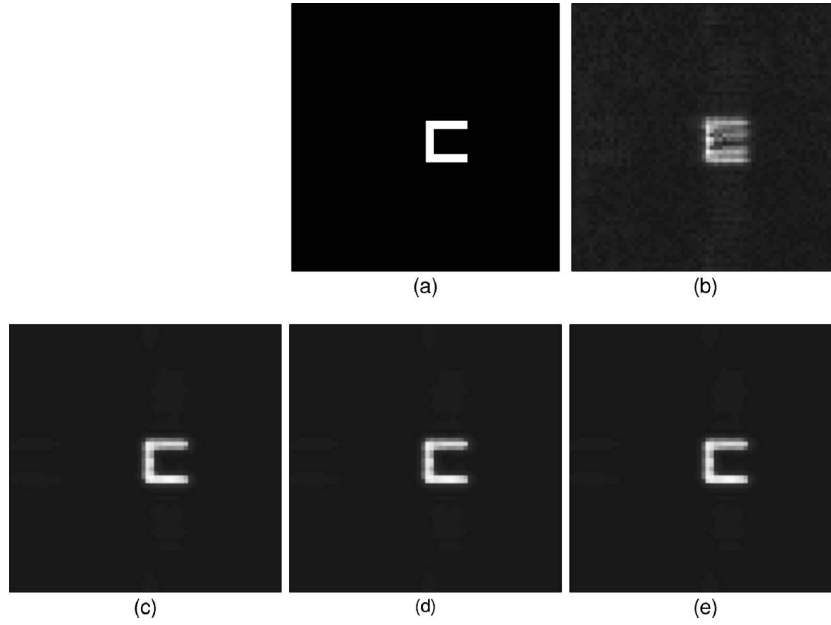
100

Fig. 5 Ultrasound imaging experiments based on measured 730-kHz data. (a) Synthetic image describing the underlying scene to be imaged. (b) Conventional reconstruction. (c),(d),(e) Reconstructions produced by the technique proposed in Sec. 2, with the following choices of regularizing functionals from Table 1, and $p=1$: (c) ψ_1, (d) ψ_2, (e) ψ_3.

in such a way to achieve noise and artifact suppression together with feature preservation in the resulting images, while taking into account the nature of the signals involved in coherent imaging. In order to efficiently solve the optimization problems formulated for coherent imaging, we extend and use half-quadratic regularization methods. Our experimental study has shown the effectiveness of this strategy in obtaining reconstructions that are superior in a number of ways to conventional coherent images. The improvements provided by these reconstructions appear to be promising for visual and automatic interpretation of the underlying scenes. One interesting direction for future work is the extension of the techniques presented in this paper to coherent imaging problems involving nonlinear observation models.

5 Appendix

5.1 *Discrete 2-D Derivative Operators*

In our method, we use smoothness constraints on a field, which require the spatial derivatives of the field. We use the horizontal and vertical first-order difference operators in approximating such derivatives. Derivatives of the field in other directions, such as the diagonals, may be used as well; however, we have found the use of horizontal and vertical derivatives sufficient. Consider a real-valued, sampled field $\mathbf{z}$, column-stacked as a vector of length $N=N_xN_y$, where N_x and N_y denote the numbers of rows and columns, respectively, in the 2-D field. We can compute first differences of this field, $\mathbf{D}_x\mathbf{z}$ and $\mathbf{D}_y\mathbf{z}$, in the horizontal and vertical directions, respectively, where the discrete derivative operators are given by

$$\mathbf{D}_x = \begin{bmatrix} -\mathbf{I} & \mathbf{I} & & \\ & \ddots & \ddots & \\ & & -\mathbf{I} & \mathbf{I} \end{bmatrix} \tag{15}$$

and

$$\mathbf{D}_y = \begin{bmatrix} \mathbf{D}_1 & & & \\ & \mathbf{D}_1 & & \\ & & \ddots & \\ & & & \mathbf{D}_1 \end{bmatrix}, \tag{16}$$

with

$$\mathbf{D}_1 = \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}. \tag{17}$$

Note that, since we take first differences between neighboring pixels, it is appropriate to have the discrete derivatives defined on the locations between the adjacent pixels. With these definitions, $\mathbf{D}_x$ then has a size of $N_y(N_x-1) \times N_xN_y$, and $\mathbf{D}_y$ has a size of $N_x(N_y-1) \times N_xN_y$. Hence, these are nonsquare operators. However, if the use of square derivative operators is desired, the preceding definitions can be augmented by derivatives defined at the boundary of the field. This may be preferred, for example, when one wants to associate each derivative to a pixel location.

We now describe two ways to compute the smoothness constraint terms, of the form $\|\mathbf{Dz}\|_p^p$, that appear in objective functionals such as that in Eq. (3). The discussion can eas-

ily be generalized to smoothness constraints with other potential functionals, such as those considered in Sec. 2.5.

The first approach is based on treating the horizontal and vertical derivatives separately when imposing a smoothness constraint. This is achieved by defining the 2-D discrete derivative operator $\mathbf{D}$ as follows:

$$\mathbf{D} \triangleq \begin{bmatrix} \mathbf{D}_x \\ \mathbf{D}_y \end{bmatrix}. \tag{18}$$

With this definition, we can write $\|\mathbf{Dz}\|_p^p$ as

$$\|\mathbf{Dz}\|_p^p = \sum_{i=1}^{M} |(\mathbf{Dz})_i|^p = \sum_{i=1}^{M_x} |(\mathbf{D}_x\mathbf{z})_i|^p + \sum_{i=1}^{M_y} |(\mathbf{D}_y\mathbf{z})_i|^p$$

$$= \|\mathbf{D}_x\mathbf{z}\|_p^p + \|\mathbf{D}_y\mathbf{z}\|_p^p, \tag{19}$$

where $M_x \triangleq N_y(N_x-1)$, $M_y \triangleq N_x(N_y-1)$, and $M = M_x + M_y$.

The second approach is based on treating the gradient at each pixel location as a two-element vector $[(\mathbf{D}_x\mathbf{z})_i \ (\mathbf{D}_y\mathbf{z})_i]^T$, composed of the horizontal and vertical gradients, and using the ℓ_2 norms of such gradients at all locations in the field for the computation of the overall ℓ_p norm:

$$\|\mathbf{Dz}\|_p^p \triangleq \sum_{i=1}^{N} [|(\mathbf{D}_x\mathbf{z})_i|^2 + |(\mathbf{D}_y\mathbf{z})_i|^2]^{p/2}. \tag{20}$$

Two things must be noted here. First, the use of a *linear* operator $\mathbf{D}$ is only conceptual in this case, because no such explicit matrix exists. Second, this approach requires a one-to-one correspondence between horizontal and vertical derivatives at each location in the scene; hence in this case we use square $(N \times N)$ derivative operators $\mathbf{D}_x, \mathbf{D}_y$.

In our method, we make use of both approaches described; however, all the mathematical expressions in the body of this paper are based on the first approach. Note that when $p=2$, the two approaches are identical, with the use of square derivative operators. To make the association between the two approaches clear, let us consider square derivative operators, and examine the first approach in this case:

$$\|\mathbf{Dz}\|_p^p = \sum_{i=1}^{N} |(\mathbf{D}_x\mathbf{z})_i|^p + \sum_{i=1}^{N} |(\mathbf{D}_y\mathbf{z})_i|^p \tag{21}$$

$$= \sum_{i=1}^{N} [|(\mathbf{D}_x\mathbf{z})_i|^p + |(\mathbf{D}_y\mathbf{z})_i|^p]. \tag{22}$$

Let us compare this expression with the second approach, given in Eq. (20). There the ℓ_2 norm of the gradient vector at each location is used in the computation of the overall ℓ_p norm. In contrast, the first approach, as shown in Eq. (22), corresponds to using an ℓ_p norm for the gradient vector $[(\mathbf{D}_x\mathbf{z})_i \ (\mathbf{D}_y\mathbf{z})_i]^T$ at each location. This association lets us compare the consequences of using the two approaches. For example, when $p < 2$, the first approach used in a smoothness constraint would favor horizontal and vertical edges over diagonal edges, more than in the second approach.

5.2 Half-Quadratic Functional for ℓ_p-Norm-Based Regularization

The objective of this subsection is to prove the relationship (5), which we repeat below, between the particular functionals $J(\mathbf{f})$ of Eq. (4) and $K(\mathbf{f}, \mathbf{b}, \mathbf{s})$ of Eq. (6):

$$\inf_{\mathbf{b}, \mathbf{s}} K(\mathbf{f}, \mathbf{b}, \mathbf{s}) = J(\mathbf{f}). \tag{23}$$

This relationship shows that $K(\mathbf{f}, \mathbf{b}, \mathbf{s})$ of Eq. (6) is a valid augmented cost functional to be used in half-quadratic regularization for the functional $J(\mathbf{f})$ of Eq. (4).

To keep the derivation simple, we consider a 1-D signal $\mathbf{f}$, rather than a 2-D field, in this subsection. The results however can easily be extended to the 2-D case. We assume the following structure for the discrete 1-D derivative operator $\mathbf{D}$:

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}, \tag{24}$$

which simply consists of two-element differences.

Let us now find $\mathbf{s}$ and $\mathbf{b}$ that minimize $K(\mathbf{f}, \mathbf{b}, \mathbf{s})$ of Eq. (6). First consider $\mathbf{s}$. The portion of $K(\mathbf{f}, \mathbf{b}, \mathbf{s})$ that depends on $\mathbf{s}$ is the following:

$$\sum_{i=1}^{M} \mathbf{b}_i |(\mathbf{DSf})_i|^2. \tag{25}$$

Based on the structures of $\mathbf{D}$ in Eq. (24) and $\mathbf{S}$ in Eq. (7), we have

$$(\mathbf{DSf})_i = -\exp(-j\mathbf{s}_i)\mathbf{f}_i + \exp(-j\mathbf{s}_{i+1})\mathbf{f}_{i+1}, \tag{26}$$

and consequently

$$|(\mathbf{DSf})_i|^2 = |\mathbf{f}_i|^2 + |\mathbf{f}_{i+1}|^2 - 2\Re[|\mathbf{f}_i| \, |\mathbf{f}_{i+1}|\exp(j\{\phi[(\mathbf{f})_i]$$
$$- \phi[(\mathbf{f})_{i+1}]\})\exp[j(\mathbf{s}_{i+1} - \mathbf{s}_i)]]. \tag{27}$$

Here $\phi[(\mathbf{f})_i]$ denotes the phase of the complex number $\mathbf{f}_i$. The sum in Eq. (25) takes its minimum value when the product inside the outermost brackets in Eq. (27) has a zero imaginary part for all i. Hence the minimizing $\mathbf{s}$ satisfies

$$\mathbf{s}_{i+1} - \mathbf{s}_i + \phi[(\mathbf{f})_i] - \phi[(\mathbf{f})_{i+1}] = 0. \tag{28}$$

We could have obtained this result by the following qualitative argument as well. We want to minimize Eq. (25), which is a weighted sum of squared norms of the differences between complex-number pairs of the form $z_i = \exp(-j\mathbf{s}_i)\mathbf{f}_i$. The variables we have for optimization are $\mathbf{s}_i$ for all i, hence we can essentially choose the phase of each complex number. Naturally, the minimum is obtained when the complex numbers z_i have identical phase, since this makes the norm of the difference between two complex numbers as small as possible. This is exactly what the condition in Eq. (28) implies: the optimum $\mathbf{s}_i$ should "rotate" $\mathbf{f}_i$ in such a way that the resulting z_i have the same phase for

all i. Note that we still have a freedom in choosing what that identical phase is. If we simply choose it to be 0, then we have the following optimal **s**:

$$\mathbf{s}_i = \phi[(\mathbf{f})_i] \ \ \forall \ i. \tag{29}$$

Note that with this **s**, we have $\mathbf{Sf} = |\mathbf{f}|$. Hence,

$$\inf_{s} K(\mathbf{f},\mathbf{b},\mathbf{s}) = \|\mathbf{y} - \mathbf{Tf}\|_2^2 + \lambda \sum_{i=1}^{M} \left[\mathbf{b}_i[|(\mathbf{D}|\mathbf{f}|)_i|^2 + \epsilon] \right.$$
$$\left. + \left(\frac{p}{2\mathbf{b}_i} \right)^{\frac{p}{2-p}} \left(1 - \frac{p}{2} \right) \right]. \tag{30}$$

Next, let us consider **b**. Differentiating the summand in Eq. (30) and setting it equal to zero, we obtain the following condition for the minimizing **b**:

$$\mathbf{b}_i = \frac{p}{2[|(\mathbf{D}|\mathbf{f}|)_i|^2 + \epsilon]^{1-p/2}}. \tag{31}$$

Substituting Eq. (31) in $K(\mathbf{f},\mathbf{b},\mathbf{s})$, we obtain the result we desire:

$$\inf_{\mathbf{b},\mathbf{s}} K(\mathbf{f},\mathbf{b},\mathbf{s}) = \|\mathbf{y} - \mathbf{Tf}\|_2^2 + \lambda \sum_{i=1}^{M} [|(\mathbf{D}|\mathbf{f}|)_i|^2 + \epsilon]^{p/2} = J(\mathbf{f}), \tag{32}$$

which shows that Eq. (5) holds for $J(\mathbf{f})$ of Eq. (4) and $K(\mathbf{f},\mathbf{b},\mathbf{s})$ of Eq. (6).

5.3 Convergence of the Algorithm in Sec. 2.3

Let us consider the sequence $K_n = K(\hat{\mathbf{f}}^{(n)}, \hat{\mathbf{b}}^{(n+1)}, \hat{\mathbf{s}}^{(n+1)})$, and show that it is convergent. From Eqs. (8) and (9), we have

$$K(\hat{\mathbf{f}}^{(n)}, \hat{\mathbf{b}}^{(n)}, \hat{\mathbf{s}}^{(n+1)}) \leq K(\hat{\mathbf{f}}^{(n)}, \hat{\mathbf{b}}^{(n)}, \hat{\mathbf{s}}^{(n)}) \ \ \forall \ n, \tag{33}$$

$$K(\hat{\mathbf{f}}^{(n)}, \hat{\mathbf{b}}^{(n+1)}, \hat{\mathbf{s}}^{(n+1)}) \leq K(\hat{\mathbf{f}}^{(n)}, \hat{\mathbf{b}}^{(n)}, \hat{\mathbf{s}}^{(n+1)}) \ \ \forall \ n, \tag{34}$$

which implies

$$K(\hat{\mathbf{f}}^{(n)}, \hat{\mathbf{b}}^{(n+1)}, \hat{\mathbf{s}}^{(n+1)}) \leq K(\hat{\mathbf{f}}^{(n)}, \hat{\mathbf{b}}^{(n)}, \hat{\mathbf{s}}^{(n)}). \tag{35}$$

Similarly, from Eq. (10), we have

$$K(\hat{\mathbf{f}}^{(n+1)}, \hat{\mathbf{b}}^{(n+1)}, \hat{\mathbf{s}}^{(n+1)}) \leq K(\hat{\mathbf{f}}^{(n)}, \hat{\mathbf{b}}^{(n+1)}, \hat{\mathbf{s}}^{(n+1)}) \ \ \forall \ n. \tag{36}$$

Now, let us consider the difference:

$$K_n - K_{n-1} = \left[K(\hat{\mathbf{f}}^{(n)}, \hat{\mathbf{b}}^{(n+1)}, \hat{\mathbf{s}}^{(n+1)}) - K(\hat{\mathbf{f}}^{(n)}, \hat{\mathbf{b}}^{(n)}, \hat{\mathbf{s}}^{(n)}) \right]$$
$$+ \left[K(\hat{\mathbf{f}}^{(n)}, \hat{\mathbf{b}}^{(n)}, \hat{\mathbf{s}}^{(n)}) - K(\hat{\mathbf{f}}^{(n-1)}, \hat{\mathbf{b}}^{(n)}, \hat{\mathbf{s}}^{(n)}) \right]. \tag{37}$$

Using Eqs. (35) and (36), we obtain

$$K_n - K_{n-1} \leq 0 \ \ \forall \ n, \tag{38}$$

which means that the sequence K_n is decreasing. Since it is bounded below and decreasing, the sequence K_n converges. Hence the algorithm in Sec. 2.3 is convergent in terms of the cost functional. A similar convergence result can be applied to the variants of this algorithm described in later sections.

References

1. J. M. Blackledge, *Quantitative Coherent Imaging*, Academic Press, San Diego, CA (1989).
2. A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Sov. Math. Dokl.* **4**, 1035–1038 (1963).
3. C. R. Vogel and M. E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Trans. Image Process.* **7**(6), 813–824 (1998).
4. P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Process.* **6**(2), 298–310 (1997).
5. D. C. Munson, Jr., and J. L. C. Sanz, "Image reconstruction from frequency-offset Fourier data," *Proc. IEEE* **72**(6), 661–669 (1984).
6. D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Natl. Acad. Sci. U.S.A.* **100**(5), 2197–2202 (2003).
7. D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(3), 367–383 (1992).
8. J. He and S. X. Pan, "Magnitude reconstruction of complex images from incomplete Fourier phase data," in *16th Annual Conf. of the IEEE Industrial Electronics Society*, Vol. **1**, pp. 357–362 (1990).
9. B. K. Jennison and J. P. Allebach, "Maximum likelihood image reconstruction from Fourier-offset data using the expectation-maximization algorithm," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vol. **4**, pp. 2597–2600 (1991).
10. M. Rabbani and B. E. A. Saleh, "Bayesian filtering of speckled images," *Opt. Commun.* **53**(3), 147–152 (1985).
11. J. F. Giovannelli, A. Herment, and G. Demoment, "A Bayesian approach to ultrasound Doppler spectral analysis," in *IEEE Ultrasonics Symp.*, pp. 1055–1058 (1993).
12. C. P. Mariadassou and B. Yegnanarayana, "Image reconstruction from noisy digital holograms," *IEE Proc. F, Radar Signal Process.* **137**(5), 351–356 (1990).
13. H. Carfantan and A. Mohammad-Djafari, "A Bayesian approach for nonlinear inverse scattering tomographic imaging," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vol. **4**, pp. 2311–2314 (1995).
14. D. J. Battle, R. P. Harrison, and M. Hedley, "Maximum entropy image reconstruction from sparsely sampled coherent field data," *IEEE Trans. Image Process.* **6**(8), 1139–1147 (1997).
15. D. J. Battle, "Maximum entropy regularisation applied to ultrasonic image reconstruction," PhD Thesis, Univ. of Sydney (1999).
16. J. M. Gorce, D. Friboulet, J. D'hooge, B. Bijnens, and I. E. Magnin, "Regularized autoregressive models for a spectral estimation scheme dedicated to medical ultrasonic radio-frequency images," in *IEEE Ultrasonics Symp.*, pp. 1461–1464 (1997).
17. P. Ciuciu, J. Idier, and J. F. Giovannelli, "Regularized estimation of mixed spectra using a circular Gibbs-Markov model," *IEEE Trans. Image Process.* **49**(10), 2202–2213 (2001).
18. O. Husby, T. Lie, T. Langø, J. Hokland, and H. Rue, "Bayesian 2-D deconvolution: a model for diffuse ultrasound scattering," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **48**(1), 121–130 (2001).
19. K. Z. Abd-Elmoniem, A.-B. M. Yousef, and Y. M. Kadah, "Real-time speckle reduction and coherence enhancement in ultrasound imaging via nonlinear anisotropic diffusion," *IEEE Trans. Biomed. Eng.* **49**(9), 997–1014 (2002).
20. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(7), 629–639 (1990).
21. A. Abubakar and P. M. van den Berg, "A multiplicative weighted l^2-norm total variation regularization for deblurring algorithms," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vol. **4**, pp. 3545–3548 (2002).
22. S. Sotthivirat and J. A. Fessler, "Penalized-likelihood image reconstruction for digital holography," *J. Opt. Soc. Am. A* **21**(5), 737–750

(2004).

23. A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, IEEE Press, New York (1988).
24. L. P. Yaroslavskii and N. S. Merzlyakov, *Methods of Digital Holography*, Consultants Bureau, New York (1980).
25. D. C. Munson, Jr., J. D. O'Brien, and W. K. Jenkins, "A tomographic formulation of spotlight-mode synthetic aperture radar," *Proc. IEEE* **71**, 917–925 (Aug. 1983).
26. C. V. Jakowatz, Jr., D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and P. A. Thompson, *Spotlight-mode Synthetic Aperture Radar: A Signal Processing Approach*, Kluwer Academic Publishers, Norwell, MA (1996).
27. J. A. Jensen, *Estimation of Blood Velocities Using Ultrasound: A Signal Processing Approach*, Cambridge University Press, Cambridge, UK (1996).
28. J. T. Ylitalo and H. Ermert, "Ultrasound synthetic aperture imaging: monostatic approach," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **41**(3), 333–339 (1994).
29. D. L. Donoho, I. M. Johnstone, J. C. Koch, and A. S. Stern, "Maximum entropy and the nearly black object," *J. R. Stat. Soc. Ser. B. Methodol.* **54**(1), 41–81 (1992).
30. G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins Univ. Press, Baltimore (1996).
31. A. H. Delaney and Y. Bresler, "Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography," *IEEE Trans. Image Process.* **7**(2), 204–221 (1998).
32. W. C. Karl, "Regularization in image restoration and reconstruction," in *Handbook of Image and Video Processing*, A. Bovik, Ed., pp. 141–160, Academic Press (2000).
33. S. Geman and D. E. McClure, "Statistical methods for tomographic image reconstruction," in *Proc. 46th Sess. Int. Statistical Institute, Bull. ISI* **52** (1987).
34. T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Med. Imaging* **8**, 194–202 (1989).
35. Boston University Acoustics Group, Medical Ultrasound Testbed (MedBED), Web page, http://www.bu.edu/paclab/censsis/medbed/
36. R. L. Lagendijk and J. Biemond, *Iterative Identification and Restoration of Images*, Kluwer, Boston, (1991).

Müjdat Çetin received his BS degree from Boğaziçi University, Istanbul, Turkey, in 1993, his MS degree from the University of Salford, Manchester, UK, in 1995, and his PhD degree from Boston University, Boston, MA, USA, in 2001, all in electrical engineering. Since 2001, he has been a member of the Laboratory for Information and Decision Systems, M.I.T., Cambridge, MA, USA, where he now holds the title of research scientist. Since September 2005 he has also been an assistant professor at Sabanci University, Istanbul, Turkey. He has served in various organizational capacities, including special session organizer, session chair, and technical program committee member, for a number of conferences, including the SPIE Conference on Algorithms for Synthetic Aperture Radar Imagery, the IEEE International Conference on Image Processing, and the IEEE International Conference on Acoustics, Speech, and Signal Processing. His research interests are in the broad areas of statistical signal and image processing, inverse problems, computer vision, and data fusion.

William C. Karl received his PhD degree in electrical engineering and computer science in 1991 from the Massachusetts Institute of Technology, Cambridge, where he also received his SM, EE, and SB degrees. He held the position of staff research scientist with the Brown-Harvard-M.I.T. Center for Intelligent Control Systems and the M.I.T. Laboratory for Information and Decision Systems from 1992 to 1994. He joined the faculty of Boston University in 1995, where he is currently professor of electrical and computer engineering. Since January 1996 he has also held a joint appointment in the Department of Biomedical Engineering at Boston University. He has been an associate editor of the *IEEE Transactions on Image Processing*. He has also served in various organizational capacities, including session organizer and chair for the 36th Asilomar Conference on Signals, Systems and Computers, special session on inverse problems in imaging; session organizer and chair for the 2000 Conference in Information Sciences and Systems, special session on medical imaging; and member of the organizing committee for the First SIAM Conference on the Life Sciences. Dr. Karl's research interests are in the areas of multidimensional signal and image processing, geometric estimation, detection, and medical signal and image processing. He is a member of the Institute of Electrical and Electronics Engineers and Sigma Xi.
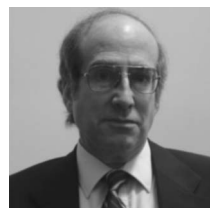
Alan S. Willsky joined the M.I.T. faculty in 1973 and is currently the Edwin Sibley Webster Professor of Electrical Engineering. He is a founder, member of the board of directors, and chief scientific consultant of Alphatech, Inc. From 1998–2002 he served as a member of the U.S. Air Force Scientific Advisory Board. He has received several awards, including the 1975 American Automatic Control Council Donald P. Eckman Award, the 1979 ASCE Alfred Noble Prize, and the 1980 IEEE Browder J. Thompson Memorial Award. Dr. Willsky has held visiting positions in England and France and various leadership positions in the IEEE Control Systems Society (which made him a Distinguished Member in 1988). He has delivered numerous keynote addresses and is coauthor of the undergraduate text *Signals and Systems*. His research interests are in the development and application of advanced methods of estimation and statistical signal and image processing. Methods he has developed have been successfully applied in a variety of applications, including failure detection, surveillance systems, biomedical signal and image processing, and remote sensing.

SPARSITY-DRIVEN SPARSE-APERTURE ULTRASOUND IMAGING

Müjdat Çetin[a,b], Emmanuel Bossy[c], Robin Cleveland[c], and W. Clem Karl[d]

[a]Faculty of Engineering and Natural Sciences, Sabancı University, İstanbul, Turkey
[b]Laboratory for Information and Decision Systems, MIT, Cambridge, MA, USA
[c]Department of Aerospace and Mechanical Engineering, Boston University, Boston, MA, USA
[d]Department of Electrical and Computer Engineering, Boston University, Boston, MA, USA

ABSTRACT

We propose an image formation algorithm for ultrasound imaging based on sparsity-driven regularization functionals. We consider data collected by synthetic transducer arrays, with the primary motivating application being nondestructive evaluation. Our framework involves the use of a physical optics-based forward model of the observation process; the formulation of an optimization problem for image formation; and the solution of that problem through efficient numerical algorithms. Our sparsity-driven, model-based approach achieves the preservation of physical features while suppressing spurious artifacts. It also provides robust reconstructions in the case of sparse observation apertures. We demonstrate the effectiveness of our imaging strategy on real ultrasound data.

1. INTRODUCTION

Nondestructive evaluation (NDE) of materials is a critical task in applications including defense, nuclear power, manufacturing, and infrastructure monitoring [1]. Through imaging, one could view the internal structure of homogeneous materials to determine the presence, severity, and characteristics of inhomogeneities, such as cracks. Ultrasound continues to be the imaging modality of choice in many NDE scenarios due to its safety, versatility, and low cost [1]. There are a number of data collection and imaging setups, and here we focus on a monostatic configuration, consisting of a non-focused transducer mechanically scanned to construct a synthetic aperture. At each mechanically scanned position, the transducer sends acoustic pulses and records the scattered waveforms. Given such data, the goal is to reconstruct a 3-D image of the material or that of a 2-D cross section. A conventional technique to reconstruct images is beamforming, which suffers from poor resolution and sidelobe artifacts. One could also consider data inversion through a pseudoinverse operation, which is very sensitive to noise in the data.

A current trend in many imaging applications is to develop and study imaging strategies for the case of *sparse apertures*, in which the data lie in a small and potentially irregular portion of what would be considered a full aperture of spatial or spectral observation points. In some applications sparse apertures emerge as a result of physical or geometric constraints in the observation scenario (e.g. we cannot place the sensor at a particular location). In other applications, such apertures are of interest, because sensing is viewed as a dear resource, and the goal is to form accurate images with as little data as possible. When data are limited and lie on an irregular grid, conventional imaging strategies suffer severely from degraded resolution and imaging artifacts. For the practical utility of such sparse-aperture sensing scenarios, advanced image formation algorithms that produce enhanced imagery facilitating visual or automatic interpretation of the underlying scenes are needed.

We propose a new approach for ultrasound imaging to produce enhanced images especially in challenging scenarios involving sparse apertures. The primary application that has motivated us is nondestructive evaluation, although the approach could be adapted to other applications as well. Our framework is based on a regularized reconstruction of the underlying reflectivity field based on the scattered ultrasound data. We use nonquadratic regularization functionals which exploit the expected sparsity of the underlying fields. In our previous work, we have applied such sparsity-driven approaches to other wave-based imaging problems such as radar imaging [2]. Such functionals enable the preservation of strong physical features (such as strong scatterers or boundaries between regions with different reflectivity properties), and have been shown to lead to superresolution. We combine such functionals with a data fidelity term based on a physical optics-based linear model of the observation process to formulate an optimization problem for image formation. We solve the resulting optimization problem using efficient numerical algorithms.

There are a number of publications which have relations to our perspective for ultrasound imaging. A Bayesian approach for the nonlinear inverse scattering problem of tomographic imaging using ultrasound probing has been proposed in [3]. In [4], maximum entropy regularization has been used for image reconstruction from sparsely sampled coherent field data. The work in [5] proposes a regularized autoregressive model for spectral estimation, with application to medical ultrasonic radio-frequency images. A statistical deconvolution technique for diffuse ultrasound scattering has been proposed in [6], where sampling techniques are used for inference. Finally the approach in a recent thesis [7], performed indepen-

dently from our work, shares some of the key ideas in this paper.

There are a number of aspects of our work that differentiate it from existing literature. A detailed comparison is beyond the scope of this paper, but some key aspects of our work include: use of ℓ_p-norms for regularization which can seamlessly handle complex-valued data; use of sparsity constraints both on the complex-valued reflectivity field as well as on the gradient of its magnitude; development and use of efficient optimization algorithms matched to the problem structure. Given the previous work by us and others on the use of these types of algorithms in other applications, the contribution of this paper is the adaptation of these ideas to the ultrasound imaging modality through the incorporation of a physics-based forward model, as well as demonstration of the effectiveness of the approach on real, sparse-aperture ultrasound data. In particular, these experiments show how the proposed approach can provide improved resolution, reduced artifacts, and robustness to aperture sparsity as compared to conventional imaging methods.

2. OBSERVATION MODEL FOR ULTRASOUND SCATTERING

The observation model we use for ultrasound scattering is based on a linearization of the scalar wave equation. We use the following Green's function to model the scattered field in space in response to a point source of excitation:

$$G(|\mathbf{r}' - \mathbf{r}|) = \frac{\exp(jk(|\mathbf{r}' - \mathbf{r}|)}{4\pi|\mathbf{r}' - \mathbf{r}|} \qquad (1)$$

where $\mathbf{r}$ and $\mathbf{r}'$ denote the source location and the observation location, respectively, in three-dimensional space, and k is the wavenumber. In this paper we consider a monostatic data acquisition scenario. In specifying the response of a scatterer to an incident field emitted by a transducer, we assume the case of impenetrable scatterers. This is reasonable for a nondestructive evaluation application since inhomogeneities such as cracks act as strong reflectors of ultrasound energy. This leads us to use the physical optics approximation in linearizing the wave equation to obtain the following observation model:

$$y(\mathbf{r}') = 2jk \int G^2(|\mathbf{r}' - \mathbf{r}|)f(\mathbf{r})d\mathbf{r} \qquad (2)$$

where $y(\cdot)$ denotes the observed data and $f(\cdot)$ denotes the underlying, unknown reflectivity field. Note that squaring the Green's function captures the two-way travel from the transducer to the target and back. Also note that the observation model in (2) involves essentially a shift invariant point spread function. We discretize this model and take into account the presence of measurement noise to obtain the following discrete observation model:

$$\mathbf{y} = \mathbf{Tf} + \mathbf{n} \qquad (3)$$

where $\mathbf{y}$ and $\mathbf{n}$ denote the measured data and the noise, respectively, at all transducer positions; $\mathbf{f}$ denotes the sampled unknown reflectivity field; and $\mathbf{T}$ is a matrix representing the observation kernel in (2). In particular, each row of $\mathbf{T}$ is associated with measurements at a particular transducer position. The entire set of transducer positions determines the nature of the aperture used in a particular experiment, and the matrix $\mathbf{T}$ carries information about the geometry and the sparsity of the aperture.

3. SPARSITY-DRIVEN IMAGING

Given the noisy observation model in (3), the imaging problem is to find an estimate of $\mathbf{f}$ based on the data $\mathbf{y}$. In general this is an ill-posed inverse problem, and its solution requires the incorporation of explicit or implicit prior information or constraints about the underlying field $\mathbf{f}$. One type of generic prior information that has recently been successfully applied in a number of imaging applications involves the *sparsity* of some aspect of the underlying field. In the context of ultrasound imaging for nondestructive evaluation, such sparsity priors could also be a valuable asset, as we expect the underlying homogeneous material to be fairly sparse in terms of both the location of inhomogeneities (e.g. cracks), as well as the boundaries between such inhomogeneities and the homogeneous background.

It has been observed that imposing sparsity directly leads to combinatorial optimization problems, but both empirical and recent theoretical results suggest that this could in practice be achieved by relaxed and tractable nonquadratic optimization formulations, based on e.g. ℓ_p-norms (see, e.g. [8]). This is the strategy we adopt in this paper. In particular we propose to find the reconstructed image $\hat{\mathbf{f}}$ as the minimizer of the following cost functional:

$$J(\mathbf{f}) = \|\mathbf{y} - \mathbf{Tf}\|_2^2 + \lambda_1 \|\mathbf{f}\|_p^p + \lambda_2 \|\nabla|\mathbf{f}|\|_p^p \qquad (4)$$

where $\|\cdot\|_p$ denotes the ℓ_p-norm ($0 < p \leq 1$), ∇ denotes a discrete approximation to the spatial gradient operator, $|\mathbf{f}|$ denotes the vector of magnitudes of the complex-valued vector $\mathbf{f}$, and λ_1, λ_2 are scalar parameters. The first term of $J(\mathbf{f})$ in (4) is a data fidelity term, while the other terms are regularizing sparsity constraints. In particular, the second term has the role of preserving strong scatterers such as cracks while suppressing artifacts (these types of constraints lead to superresolution). The third term has the role of smoothing homogeneous regions while preserving sharp transitions, such as those between cracks and the background. The relative magnitudes of the scalar parameters λ_1 and λ_2 determine the emphasis on each term. In our experimental work, we use the second term as the dominant one, and use values of p around 1. We solve the optimization problem in (4) by adapting efficient iterative algorithms we have developed in our previous work [2] to the ultrasound imaging application.

4. EXPERIMENTS

We present the results of imaging experiments based on data collected at the Large Ultrasound Test Facility (LUTF) [9] at Boston University.

4.1. Data Collection

In our experiments we use a tank full of water as the homogeneous material in which waves propagate. We insert an aluminum object inside this homogeneous medium as the inhomogeneity. The objective of the imaging experiments is
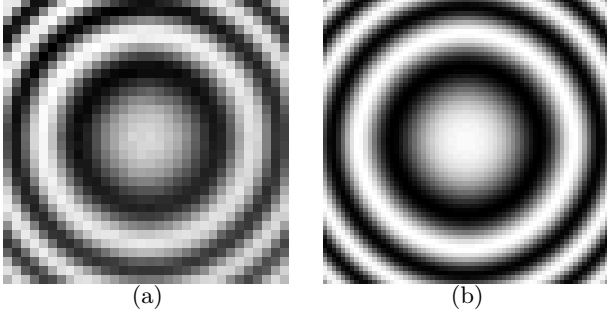
Fig. 1. The point spread function (PSF) of the data collection system at 300 kHz. (a) Measured using a 1 mm diameter spherical scatterer. (b) Theoretical model. (Real parts shown.)

Fig. 2. (a) Location and shape of the cross-section of the aluminum object to be imaged. (b) Real part of the data scattered by the object at 300 kHz.

to reconstruct a 2-D cross section of this object. We use a monostatic arrangement in which a broadband single-element unfocused transducer is mechanically moved on a 64×64 grid of locations (covering a square with a side of 76.8 mm.) at the top of the tank to send and receive acoustic waveforms. We place the object to be imaged at a depth of 175 mm, and time-gate the reflected signals to isolate the response from that depth. The transducer emits a broadband signal, whose two most significant peaks are at 730 kHz and 300 kHz (with the corresponding wavelengths of 2 mm and 5 mm). We transform the time-gated received signal to the frequency domain and extract the response at these two frequencies. Although our framework is suitable for processing multi-channel data, in this paper we focus on processing single-channel data at each of these two frequencies.

In order to experimentally estimate the impulse response of the system to test the validity of the theoretical model described in Section 2, we have first collected data from a spherical aluminum scatterer of 1 mm diameter. Real part of the data measured at 300 kHz through the full 64×64 aperture is displayed in Fig. 1(a). Real part of the point spread function based on the theoretical model in (2) is shown in Fig. 1(b), which is in very well agreement with Fig. 1(a). We use this theoretical model to construct the operator $\mathbf{T}$ in our experiments.

In Fig. 2(a) we illustrate the shape and the location of the U-shaped cross section of the aluminum object to be imaged (our use of this shape is inspired by the experiments in [4]). The length of each side is 12 mm, and the thickness is 2.4 mm. Real part of the measured full-aperture data at 300 kHz in the presence of this object in the tank is shown in Fig. 2(b).

4.2. Results

We first present the results of full-aperture imaging experiments based on the type of scattered data shown in Fig. 2(b). Fig. 3(a) and (b) show the results of two conventional imaging strategies, beamforming and regularized pseudoinverse [10], respectively, at 730 kHz and 300 kHz. At 730 kHz beamforming produces a good reconstruction, however when we reduce the operating frequency to 300 kHz significant resolution loss occurs. Low-frequency operation is of interest because acoustic waves suffer from more attenuation as the frequency is increased. The regularized pseudoinverse approach aims to
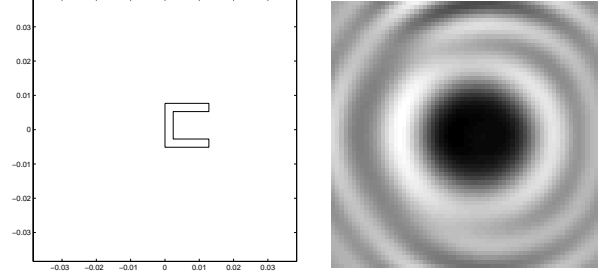
improve upon the pseudoinverse operation (which is unstable in the presence of measurement noise, and consequently not considered here) by solving the problem iteratively and providing regularization through early stopping. Although this gets rid of severe artifacts, the resulting images shown in Fig. 3(b) still do not exhibit the shape of the inhomogeneity very accurately. The reconstructions obtained using our proposed approach are shown in Fig. 3(c) and provide much more accurate images of the U-shaped object, even at the low operating frequency of 300 kHz. In our experiments, we use $p = 1$, and $\lambda_1 \gg \lambda_2$.[1] This relative choice of λ_1 and λ_2 indicates our emphasis on preserving and sharpening the strong scattering from inhomogeneities in the scene while suppressing background artifacts.

Next we consider a sparse aperture, in particular the star-shaped synthetic aperture shown in Fig 4. (Note that the full aperture used in the previous experiments was based on measurements on the 64×64 square region in Fig 4.) The number of data collection points in this sparse aperture is only 6% of the full aperture considered in the previous experiments. The imaging results are shown in Fig. 5. The conventional images shown in Fig. 4(a) and (b) suffer from insufficient resolvability of fine features and sidelobe artifacts caused by the sparsity of the aperture, making it difficult to infer the shape of the inhomogeneity. Our approach is able to suppress such artifacts and recover the shape as shown in Fig 4(c). These results illustrate the robustness of our strategy to data limitations due to the sparsity of the aperture. The experiments we have conducted are based on data carefully collected in a controlled environment, and hence represent a high-SNR scenario. We also expect our imaging strategy to provide improved robustness in low-SNR data collection scenarios.

5. CONCLUSION

We have proposed and demonstrated a sparsity-driven image formation approach for ultrasound imaging with application to nondestructive evaluation. Attractive characteristics of the proposed technique include improved resolvability of fine features, suppression of artifacts, and robustness to the sparsity of the observation aperture. Based on the initial work presented in this paper, a number of directions emerge as potential research topics. First, although the study in this

---

[1]We do not specify the actual values as they depend on the scaling of the data in a particular experiment, and hence are not very informative.
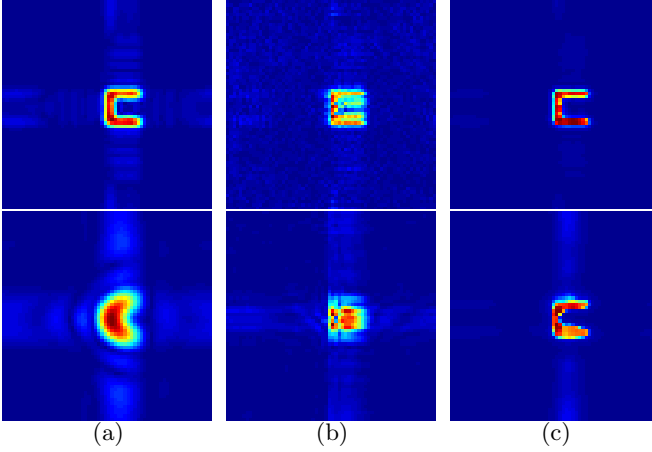
Fig. 3. Reconstructed images from full-aperture data. (Magnitudes shown.) Top: 730 kHz. Bottom: 300 kHz. (a) Beamforming. (b) Regularized pseudoinverse. (c) Proposed method.

Fig. 5. Reconstructed images from sparse-aperture data. (Magnitudes shown.) Top: 730 kHz. Bottom: 300 kHz. (a) Beamforming. (b) Regularized pseudoinverse. (c) Proposed method.
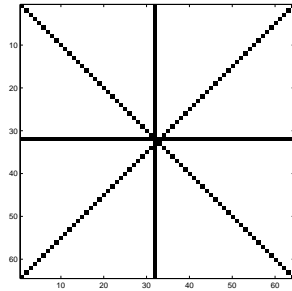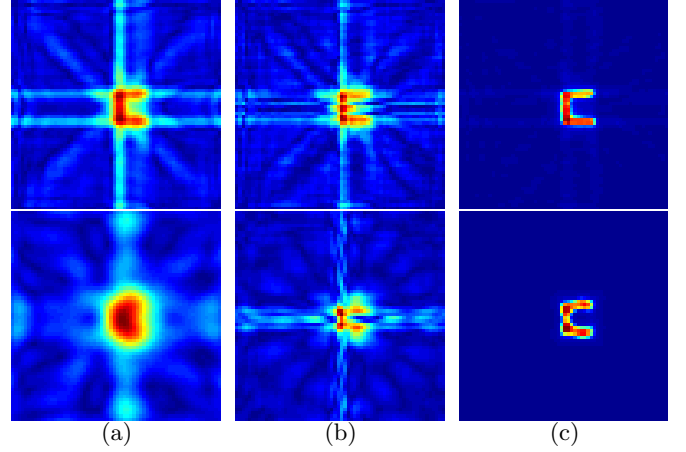

Fig. 4. Transducer positions used to construct a sparse aperture. Relative to the full-aperture used, this aperture is 6% filled.

paper was limited to monostatic, single-channel data, extension of the developed framework to the multistatic case, as well as to the processing of multi-channel data is straightforward. Our work could also be used for forming 3-D images. For the experimental setup considered in this paper, a linear observation model based on a single-scattering assumption was reasonable, however it might be of interest to generalize the framework to the case of multiple scattering and nonlinear models. It is also worthwhile to characterize the behavior of the proposed approach as the problem becomes more challenging (e.g. through the scene content, frequency of operation, sparsity of the aperture, etc.), and understand how the performance of the proposed approach degrades. Finally, although nondestructive evaluation was the motivating application here, it is of interest to adapt and apply this technique on other ultrasound applications, the most notable one being medical imaging.

6. REFERENCES

[1] P. J. Shull, Ed., *Nondestructive Evaluation: Theory, Techniques, and Applications*, CRC Press, 2002.

[2] M. Çetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Trans. Image Processing*, vol. 10, no. 4, pp. 623–631, Apr. 2001.

[3] H. Carfantan and A. Mohammad-Djafari, "A Bayesian approach for nonlinear inverse scattering tomographic imaging," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1995, vol. 4, pp. 2311–2314.

[4] D. J. Battle, R. P. Harrison, and M. Hedley, "Maximum entropy image reconstruction from sparsely sampled coherent field data," *IEEE Trans. Image Processing*, vol. 6, no. 8, pp. 1139–1147, Aug. 1997.

[5] J. M. Gorce, D. Friboulet, J. D'hooge, B. Bijnens, and I. E. Magnin, "Regularized autoregressive models for a spectral estimation scheme dedicated to medical ultrasonic radio-frequency images," in *IEEE Ultrasonics Symposium*, 1997, pp. 1461–1464.

[6] O. Husby, T. Lie, T. Langø, J. Hokland, and H. Rue, "Bayesian 2-D deconvolution: A model for diffuse ultrasound scaterring," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 48, no. 1, pp. 121–130, Jan. 2001.

[7] R. J. L. Montero, "Pulse-echo image formation using nonquadratic regularization with speckle-based images," M.S. thesis, Univ. of Illinois Urbana-Champaign, 2005.

[8] D. M. Malioutov, M. Çetin, and A. S. Willsky, "Optimal sparse representations in general overcomplete bases," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, Canada, May 2004.

[9] "Boston University Acoustics Group, Medical Ultrasound Testbed (MedBED) Web Page: http://www.bu.edu/paclab/censsis/medbed/," .

[10] R. L. Lagendijk and J. Biemond, *Iterative Identification and Restoration of Images*, Kluwer, Boston, MA, 1991.

Joint space aspect reconstruction of wide-angle SAR exploiting sparsity

Ivana Stojanovic[a], Mujdat Cetin[b], William C. Karl[c]

[a,c]ECE Department, Boston University, Boston, USA
[b]Sabanci University, Istanbul, Turkey

ABSTRACT

In this paper we present an algorithm for wide-angle synthetic aperture radar (SAR) image formation. Reconstruction of wide-angle SAR holds a promise of higher resolution and better information about a scene, but it also poses a number of challenges when compared to the traditional narrow-angle SAR. Most prominently, the isotropic point scattering model is no longer valid. We present an algorithm capable of producing high resolution reflectivity maps in both space and aspect, thus accounting for the anisotropic scattering behavior of targets. We pose the problem as a non-parametric three-dimensional inversion problem, with two constraints: magnitudes of the backscattered power are highly correlated across closely spaced look angles and the backscattered power originates from a small set of point scatterers. This approach considers jointly all scatterers in the scene across all azimuths, and exploits the sparsity of the underlying scattering field. We implement the algorithm and present reconstruction results on realistic data obtained from the XPatch Backhoe dataset.

Keywords: SAR, wide-angle, sparse measurements, edge-preserving regularization

1. INTRODUCTION

Wide-angle SAR (WSAR), where radar returns are collected over a large range of angles, holds the promise of increased spatial resolution. However, in collecting data over such a large angular range a number of the assumptions used in standard, narrow-angle SAR are violated. In particular, the common assumption that target reflectivity is only a function of spatial location, and not aspect, is no longer a good approximation to reality. Over large angular extents the energy reflected by targets is, in general, not uniform and most targets exhibit only limited scattering persistence.[1]

As a result, standard Fourier-based SAR image formation algorithms, such as the polar-format algorithm, perform poorly. The resulting imagery produced by these methods have limited resolution and display confounding artifacts.[2] Overall, these methods fail to completely realize the potential of WSAR.

Wide angle SAR reconstruction has been addressed in several papers. In one work, WSAR is approached as a collection of multiple overlapping 20° sub-apertures and reflectivity functions in each sub-aperture are independently reconstruct via the conventional polar-format algorithm or point-enhanced l_p norm regularization.[2] Alternatively, the problem is approached as a sparse, inverse problem over an overcomplete dictionary, with a dictionary element representing a prescribed reflectivity signature of a spatial pixel along the azimuth direction.[3]

In this paper, we also consider a spotlight synthetic aperture radar system (SAR)[4] with collocated transmitter and receiver operating in a monostatic configuration over a large angular range. Similar to the overcomplete dictionary approach,[3] we explicitly model the anisotropy of the target scattering behavior and estimate the angle-dependent scattering behavior at each scatter location. In contrast to previous approaches, however, we approach the problem as a direct, non-parametric reconstruction of the entire three-dimensional angle-dependent scattering field. We exploit the correlations in target reflectivity in aspect and the spatial sparsity of target scattering by including priors on this behavior in the reconstruction process. This approach does not require detailed prior knowledge of scatter type, yet can successfully focus information in the data. In addition, this approach provides robustness to data loss, allowing preservation of image quality from reduced data.

The rest of the paper is organized as follows. In Section 2 we outline the basic spotlight SAR scattering physics and present the anisotropic forward scattering model we use. Section 3 outlines our inverse problem formulation and finally Section 4 gives image reconstruction results obtained by the algorithm.

2. FORWARD MODEL

Typical assumption for narrow synthetic apertures is that the reflectivity of a given spatial differential area is isotropic. While this is a reasonable assumption for narrow apertures of a few degrees, most of the scene's scatterers exhibit anisotropic response when viewed over large aspects. In contrast to isotropic scattering where the reflectivity function is a function of spatial variables (x_p, y_p), in the general case, the reflectivity is additionally dependent on the aspect angle. A backscattered signal $r_{(x_p, y_p)}(t, \theta)$ of a spatial differential area centered at (x_p, y_p) to a pulse $\gamma(t)$ transmitted at time t, with the aircraft at an aspect θ is a delayed transmitted pulse modulated by the area's anisotropic reflectivity function $s(x_p, y_p, \theta)$. Mathematically, the backscattered signal is described by the following equation:

$$r_{(x_p, y_p)}(t, \theta) = \Re \left\{ A(x_p, y_p, \theta) s(x_p, y_p; \theta) \gamma \left(t - 2 \frac{R_p(\theta)}{c} \right) \right\} dx dy,$$

where $R_p(\theta)$ is the distance from the differential area $dxdy$ to the aircraft location at the aspect θ. The factor $A(x_p, y_p, \theta)$ accounts for propagation attenuation, transmitter and receiver antenna beam patterns, etc. This factor can be safely ignored, i.e. assumed to be a constant, when the scene extent is much smaller compared to the aircraft's stand-off range and when transmit and receive antenna beampatterns are omnidirectional. Again, typical isotropic point scattering assumption is relaxed in order to account for limited reflectivity persistence over wide aspect angles.

Now, to characterize a return from a realistic complex scene, a typical set of operating assumptions are put in place. When the impinging signal wavelength is small relative to the target extent, the overall response of a complex scene is well approximated as a superposition of a set of the scene's differential scatterers. Under the single-scattering (Born) approximation there is no interaction of scene components. Assuming that the transmitted waveform is a chirp pulse $\gamma(t) = e^{j(2\pi f_c t + \alpha t^2)}$ with center frequency f_c and rate α limited to time $-\frac{T}{2} \leq t \leq \frac{T}{2}$, the received signal, after pre-processing steps of downconversion and matched filtering, is as follows:

$$r(t, \theta) = \int \int_{x^2 + y^2 \leq L} s(x, y; \theta) e^{-j\Omega(t)(x \cos(\theta) + y \sin(\theta))} dx dy, \tag{1}$$

with the spatial frequency variable $\Omega(t) = \frac{2}{c}(2\pi f_c + 2\alpha(t - \frac{2R_c}{c}))$. In the discrete world, the backscattered signal collected at discrete look angles θ_s is sampled at times t_s to allow digital signal processing. In the time interval of interest the spatial frequency $\Omega(t)$ varies in the range $\left(\frac{2\pi}{c} f_c - \frac{\alpha T}{2}, \frac{2\pi}{c} f_c + \frac{\alpha T}{2} \right)$. Typically, the time sampling points t_s are chosen such that spatial discrete frequencies $\Omega(t_s) = \frac{4\pi}{c} f_s$ cover the whole range uniformly. Assuming that the scene under surveillance consists of multitude of point scatterers at locations (x_p, y_p), the
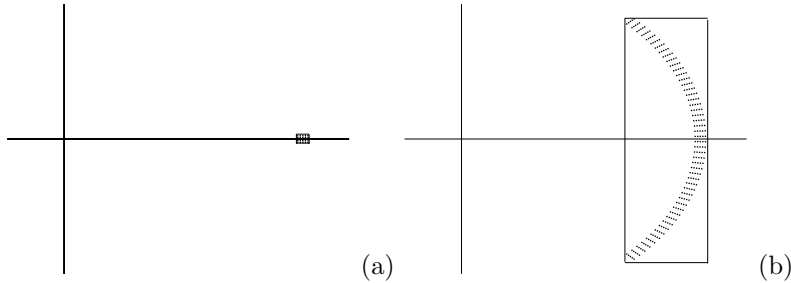

Figure 1. SAR spatial frequency support region at the center frequency $f_c = 10$ GHz: (a) narrow-band, narrow-angle case ($B = 0.5$GHz, $\Delta\theta = 5^o$) and (b) narrow-band, wide-angle ($B = 0.5$GHz, $\Delta\theta = 90^o$).

received signal can be written in discrete form as:

$$r(f_s, \theta_s) = \sum_p s(x_p, y_p; \theta_s) e^{-j \frac{4\pi f_s}{c}(x_p \cos(\theta_s) + y_p \sin(\theta_s))},$$

at a discrete frequency f_s within bandwidth B and at a discrete aspect θ_s within an aperture of the extent $\Delta\theta$, with $r(f_s, \theta_s)$ being commonly referred to as phase history. Note that the point scatterer model derived above is the discrete approximation of the continuous superposition principle that relates phase history data to continuous reflectivity field through the integral Equation 1.

Contrast between the spatial frequency support of the narrow-angle and wide-angle data collection is presented in Figure 1. Due to the circular arch shape of the spatial frequency support, traditional polar format algorithm is expected to perform poorly in the wide-angle collection scenario. Wide-angle problem is ill-posed, and direct inversion techniques result in a number of artifacts. In the following section we outline our approach that aims at joint space-aspect reconstruction of a scene viewed from wide aspect angles.

3. IMAGE FORMATION

We first define a wide angle SAR image as a set of aspect dependent spatial images. Due to the dependence of the reflectivity response on the aspect of an impinging electromagnetic wave, there exists a reflectivity map of a scene at each aspect. Assume that the ground scene is interrogated and reconstructed at a number of different aspects, I. Denote the set of time observations at the aspect θ_i as $\mathbf{r}_{\theta_i}$ and denote the spatial reflectivity field at the aspect θ_i, as $\mathbf{s}_{\theta_i}$. In the discrete representation, at each aspect angle, Equation 1 reduces to a linear system of equations of the form $\mathbf{r}_{\theta_i} = \mathbf{\Phi}_{\theta_i} \mathbf{s}_{\theta_i}$, where $\mathbf{\Phi}_{\theta_i}$ is the discrete representation of the SAR forward operator. Overall we can write:

$$\begin{bmatrix} \mathbf{r}_{\theta_1} \\ \mathbf{r}_{\theta_2} \\ \vdots \\ \mathbf{r}_{\theta_I} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_{\theta_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}_{\theta_2} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{\Phi}_{\theta_I} \end{bmatrix} \begin{bmatrix} \mathbf{s}_{\theta_1} \\ \mathbf{s}_{\theta_2} \\ \vdots \\ \mathbf{s}_{\theta_I} \end{bmatrix} + \mathbf{z}. \tag{2}$$

We can represent this relationship compactly as follows:

$$\mathbf{r} = \mathbf{\Phi}\mathbf{s} + \mathbf{z} \tag{3}$$

where $\mathbf{z}$ is a random unknown vector modeling additive system noise, as well as any model mismatch errors.

Note that two comments are in order for the above set of equations. First, the equations represent in essence a set of I independent systems of linear equations. Thus, hopes of any joint processing can not come from the forward observation model, but rather from some sort of prior information that we have about the unknown reflectivity field that we seek to reconstruct. Second, each individual problem $\mathbf{r}_{\theta_i} = \mathbf{\Phi}_{\theta_i} \mathbf{s}_{\theta_i}$ is ill-conditioned. The ill-conditioned discrete problems pose several issues in their own right, and typically some sort of prior information is utilized to aid solution stabilization and potentially reduce non-observability of components that lie in the null space of the forward operator.[5]

Under the point scattering assumption, the spatial reflectivity field at aspect i, $\mathbf{s}_{\theta_i}$, is well modeled as a spatially sparse set of reflectivity centers. Additionally, each point scatterer has a limited persistence over azimuth, but within its persistence there exist a high correlation between scatterer magnitude responses to excitations at closely spaced observation aspects. Combining these observations together, the reflectivity image magnitudes $|\mathbf{s}_{\theta_i}|$ at discrete aspects $i = \{1, \dots, I\}$ should be highly correlated, and yet allow for abrupt changes in reflectivity on a subset of scatterers. Thus, in our reconstruction algorithm we seek to impose smoothness on the point scatterer's response in the azimuth direction, and sparsity across point scatterers in the spatial domain.

To form an image we take a cost or energy minimization approach, wherein we combine the physical observation model in (2) with a term capturing prior information:

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} J_{data}(\mathbf{r}, \mathbf{s}) + J_{prior}(\mathbf{s}). \tag{4}$$

For the data-fidelity term $J_{data}(\mathbf{r}, \mathbf{s})$ we use the standard least square penalty, $J_{data}(\mathbf{r}, \mathbf{s}) = \|\mathbf{r} - \mathbf{\Phi}\mathbf{s}\|_2^2$.

We can capture the correlation in azimuth by penalizing the p-norm of the change in scattering magnitude at each pixel from angle to angle. We can capture the spatial sparsity of scatterers by penalizing the q-norm of

the total energy across aspect at each pixel[6,7] . Denote the total number of pixels in a spatial image as N. We use the following functional for the prior penalty term reflecting these insights:

$$
\begin{aligned}
J_{prior}(\mathbf{s}) \quad = \quad & \alpha \sum_{n=1}^{N} \sum_{i=1}^{I-1} ||s(x_n, y_n, \theta_{i+1})| - |s(x_n, y_n, \theta_i)||^p \\
& + \beta \sum_{n=1}^{N} \left[\sqrt{\sum_{i=1}^{I} |s(x_n, y_n, \theta_i)|^2} \right]^q .
\end{aligned}
\tag{5}
$$

Typically, we choose $p < 1$ and $q < 1$ to achieve desired sparsity[6,7] .

Note that both regularization terms are applied explicitly to magnitudes $|\mathbf{s}|$ of the complex reflectivity field $\mathbf{s}$. The second regularization term involves a l_2-norm computation which is naturally defined in terms of the magnitudes of the complex field. The first term is also expressed as the function of the field's magnitudes since it has been observed that the backscatter power is very similar across closely spaced look angles. Thus, the regularizing functional $J_{prior}(\mathbf{s})$ is non-linear function of real and imaginary parts of the field.

A solution to the inversion problem is obtained by minimizing the cost function of Equation 4. For the case when $p = q = 1$ the problem is convex and there exists a global, unique, solution. The minimization problem is in fact a second-order cone problem, that can be effectively solved by commercially available solvers. In the case when $p < 1$, $q < 1$, the convexity is lost and no local optimization algorithm can guarantee that it reaches the global minimum. Optimal sparsity is reached for $p = 0$, but the problem is then NP hard and prohibitively expensive for even moderate problem sizes. For $p,q < 1$ we use an iterative quasi-Newton method that is shown to work well on this class of problems.[8] An iterative algorithm used to find a minimizer of the cost function in Equation 4 is given in Appendix.

The computational complexity of the optimization problem in Equation 4 grows with the number of observation/reconstruction aspects. However, there is an inherit flexibility in the problem formulation, which allows for decoupling of the phase history collection aspects and the spatial field reconstruction aspects. This decoupling is carried through by mapping several azimuth returns to one spatial image. In other words, the anisotropic scattering assumption is relaxed to the isotropic within the small sub-aperture. Assume that $\{\theta_1, \ldots \theta_I\}$ now represent reconstruction aspects. At the reconstruction angle θ_i we now collect K azimuth returns, i.e. $\{\theta_i^1, \ldots \theta_i^K\}$. Thus,

$$
\mathbf{r}_{\theta_i} = \begin{bmatrix} \mathbf{r}_{\theta_i^1} \\ \vdots \\ \mathbf{r}_{\theta_i^K} \end{bmatrix}, \quad \mathbf{\Phi}_{\theta_i} = \begin{bmatrix} \mathbf{\Phi}_{\theta_i^1} \\ \vdots \\ \mathbf{\Phi}_{\theta_i^K} \end{bmatrix},
\tag{6}
$$

where $\mathbf{r}_{\theta_i^k}$ and $\mathbf{\Phi}_{\theta_i^k}$ are the discrete returned signal and the discrete forward SAR operator at the observation angle θ_i^k, respectively. The formality of the reconstruction algorithm in Equation 4 carries through unchanged with the new meaning assigned to $\mathbf{r}, \mathbf{\Phi}$ and $\mathbf{s}$.

Reduction of the number of reconstructed images poses a trade-off in between the computational complexity and the problem ill-conditioning on one side and a possible model mismatch on the other side. By assigning a small sub-aperture to each image, the degree of the ill-conditioning of a subproblem $\mathbf{r}_{\theta_i} = \mathbf{\Phi}_{\theta_i} \mathbf{s}_{\theta_i}$ is reduced by simply reducing a ratio of a number of measurements to a number unknowns. However, sub-aperture size should be chosen carefully to reduce the model mismatch. From empirical data, researchers[1] point out that the response remains isotropic, or approximately constant for aspect angles as large as 20°. Thus one could apply isotropic scattering on angular widths of a few degrees without considerably compromising the accuracy of the model.

We emphasize the spatial geometry of the data collection, as well as aspect angles at which the spatial reflectivity fields are being reconstructed on Figure 2. This figure shows a target in the coordinate center and the aircraft's circular trajectory at a large stand-off range, with phase history returns over small sub-apertures tied to one spatial image.
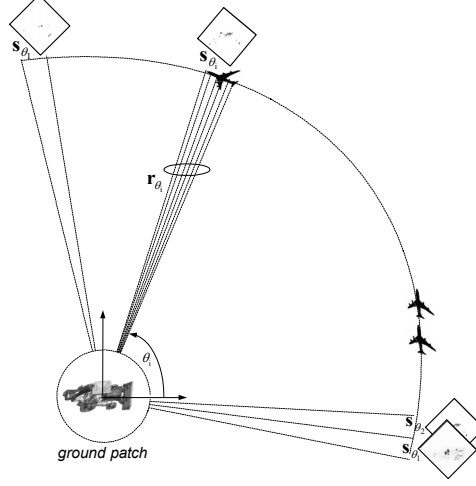
Figure 2. Wide-angle SAR data collection and reflectivity reconstruction geometry - the aircraft transmits pulses at the ground patch from a circular trajectory and reflectivity fields of the ground patch are reconstructed at a discrete set of aspects.

4. ALGORITHM ANALYSIS AND NUMERICAL SIMULATIONS

In this section we first analyze our algorithm on a synthetic data in order to derive some performance measures impossible to obtain without knowledge of the ground truth. In the second part of this section we show reflectivity field reconstructions obtained by applying the algorithm to the Backhoe dataset, generated with the XPatch simulator.[9] In both cases, we contrast joint reconstruction with independent, point enhanced processing,[2] obtained by minimizing the cost function $J(\mathbf{s}_{\theta_i}) = \|\mathbf{r}_{\theta_i} - \boldsymbol{\Phi}_{\theta_i}\mathbf{s}_{\theta_i}\|_2^2 + \beta\|\|\mathbf{s}_{\theta_i}\|\|_q^q, \forall i \in \{1, \ldots, I\}$, where $\mathbf{r}_{\theta_i}$ and $\boldsymbol{\Phi}_{\theta_i}$ are defined as before.

4.1 Performance metrics

In this section we outline performance measures used on experiments in a controlled environment to verify reconstruction abilities of our algorithm. Namely, we contrast the joint reconstruction to the point-enhanced and an ideal reconstruction (to be defined below) at a set of different signal-to-noise ratios. We compare quality of reconstruction in terms of two performance measures: relative mean squared error ($RMSE$) and percentage of correctly identified support. We introduce the ideal reconstruction to obtain a lower bound on relative mean squared error.

The relative mean squared error is defined as $RMSE = \frac{\|\widehat{\mathbf{s}} - \mathbf{s}_0\|_2^2}{\|\mathbf{s}_0\|_2^2}$, where $\widehat{\mathbf{s}}$ is a solution to an optimization problem, either joint or independent reconstruction, and $\mathbf{s}_0$ is the true underlying object. The second performance measure is a discrepancy between the ground truth support set $T = \text{supp}\{\mathbf{s}_0\}$ and a support set of a reconstructed image $\widehat{T}$. The support set $\widehat{T}$ differs form the true support set T two ways: an algorithm can introduce spurious pixel responses outside of the support set T (false alarms), or it can miss to identify a pixel in the support set T (missed detection). Due to the presence of noise, the set $\widehat{T}$ actually spreads across the whole spatial image. In order to bound the set $\widehat{T}$, one would need to threshold pixel's magnitudes below $\gamma\sigma$ to zero. The parameter γ defines a propagation of the input noise with standard deviation σ into the solution through an optimization algorithm. Instead of following this route further, we resort to a simple measure of the percentage of $|T|$ largest components of the solution $\widehat{\mathbf{s}}$ that belong to the set T.

As a baseline for $RMSE$ comparison, we define the ideal reconstruction as a reconstruction obtained by an algorithm that assumes that a spatial support is available and known at the receiver through an oracle. With the oracle help, one can *a priori* set all pixel values outside the signal support to zero and hand-pick the columns of the operator $\boldsymbol{\Phi}$ that correspond to pixels carrying the signal. The system reduces to

$$\mathbf{r} = \boldsymbol{\Phi}_T\mathbf{s}_T + \mathbf{z}$$

113

where $\mathbf{\Phi}_T$ is equivalent to the original matrix $\mathbf{\Phi}$ with appropriately pruned columns. The new signal $\mathbf{s}_T$ has the dimension $|T|$, much smaller then the original signal $\mathbf{s}$ dimension N. We also assume that the size of the measurement vector $\mathbf{r}$ is M, such that $M > |T|$. In other words, with the oracle help the problem of tackling the ill-posed inverse problem becomes a classical problem of parameter estimation in Gaussian white noise. The optimal maximum likelihood solution is equivalent to a least squares solution given by:

$$\widehat{\mathbf{s}}_T = (\mathbf{\Phi}_T^* \mathbf{\Phi}_T)^{-1} \mathbf{\Phi}_T^* \mathbf{r}.$$

Its expected mean square error is given by the formula

$$\mathbb{E}\|\widehat{\mathbf{s}}_T - \mathbf{s}_T\|_2^2 = \mathbb{E}\|(\mathbf{\Phi}_T^* \mathbf{\Phi}_T)^{-1} \mathbf{\Phi}_T^* \mathbf{z}\|_2^2 = \sigma^2 Tr((\mathbf{\Phi}_T^* \mathbf{\Phi}_T)^{-1}), \tag{7}$$

where σ^2 is noise variance and the corresponding $RMSE$ is readily derived.

Clearly the error achieved by the ideal reconstruction is a function of the matrix $\mathbf{\Phi}_T$ which in turn depends on the signal itself and number of parameters at which the system operates. Most notably, it depends on the number of measurements per image as well as the width of the viewing aperture corresponding to one image. Additionally, it depends on the distance between spatial pixels, i.e. resolution. Note that although we call this reconstruction ideal because of the oracle assistance, this reconstruction is not optimal in the sense of an achievable minimal mean square error. For the optimal reconstruction one should not assume that estimating on the true support of $\mathbf{s}_0$ achieves the minimal error. This follows by simply noting that if k-th component of the unknown $\mathbf{s}_0$ is such that its response is buried in the noise $|(\mathbf{s}_0)_k| < \gamma\sigma$, the smaller mean square error would be achieved by simply not estimating $(\mathbf{s}_0)_k$, i.e. by setting it to zero.[10] However, the optimal approach quickly becomes computationally intractable since it requires finding a least squares solution to $\mathbf{r} = \mathbf{\Phi}_T' \mathbf{s}_T' + \mathbf{z}$ for each set T' with support $T' \subset T$. The optimal reconstruction is then achieved by a LS estimator among the set of LS estimators that has the minimal relative mean square error.

4.2 Synthetic Example

The synthetic example is described as follows. We assume that a scene consists of a set of anisotropic point scatterers, i.e. a set of point scatterers reflecting non-uniformly over different aspect angles. In particular, we construct a synthetic example, pertinent to wide-angle SAR, where we are interested in uncovering a set of lexicographically ordered sparse images with two properties. First, the spatial support of any two consecutive images in the set is highly correlated and second, responses at active pixel locations across the whole set of images have limited persistence. We model an azimuth response of each active pixel as a first order Markov chain with two states: zero response and non-zero response state. Non-zero response state is modeled as a first order autoregressive process. Note that a tacit and important assumption in this study is that each image is sparse. Typical spatial 16×16 pixel reflectivity images are shown in Figure 3. Sparsity of the ground truth image is 5%. Each image corresponds to a sub-aperture of 1°. In each sub-aperture, chirp pulses interrogate the scene with 8 viewing angles and 16 frequencies over 500MHz bandwidth. Thus, we seek to uncover 256 unknown pixel responses of each image with 128 measurements per spatial image. Pixel range and cross-range resolutions are set to 0.3m. From the system parameters the predicted range and cross-range resolutions are 0.3m and 0.85m, respectively. We run the optimization algorithm in both joint processing and independent, point enhanced mode. Regularization parameters are optimized in each mode independently.

In Figure 4 we compare the performance when the set of 20 spatial images is reconstructed by joint processing of all images to the reconstruction of independent-point enhanced processing. In the point-enhanced reconstruction case, the $RMSE$ is calculated by concatenating all separately reconstructed spatial images at different aspects into one vector and applying the $RMSE$ formula. $RMSE$s achieved by these approaches as a function of signal to noise ratio is given in Figure 4(a). Theoretically predicted performance with the oracle assistance is plotted as the baseline for comparison. These results indicate that joint processing of spatial images that have highly correlated spatial support considerably reduces the error over the independent point-enhanced processing and significantly reduces the gap to ideal reconstruction.

An evaluation of the techniques in terms of correctly identified support is given in Figure 4(b). These results indicates that joint processing achieves better noise suppression. This point is further exemplified in Figure 5
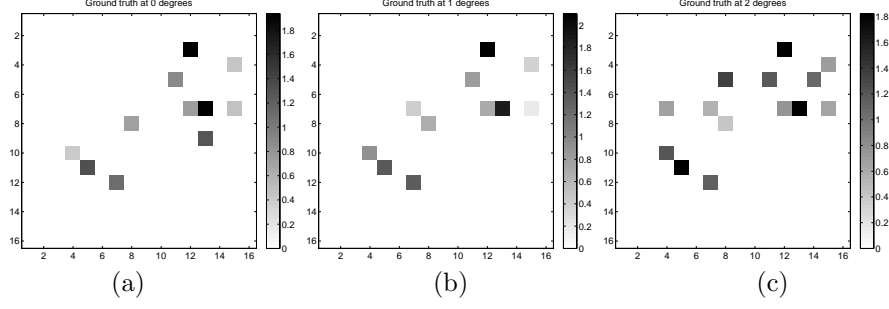
Figure 3. Typical ground truth images at three consecutive aspect angles: (a) 0°, (b) 1° and (c) 2°.
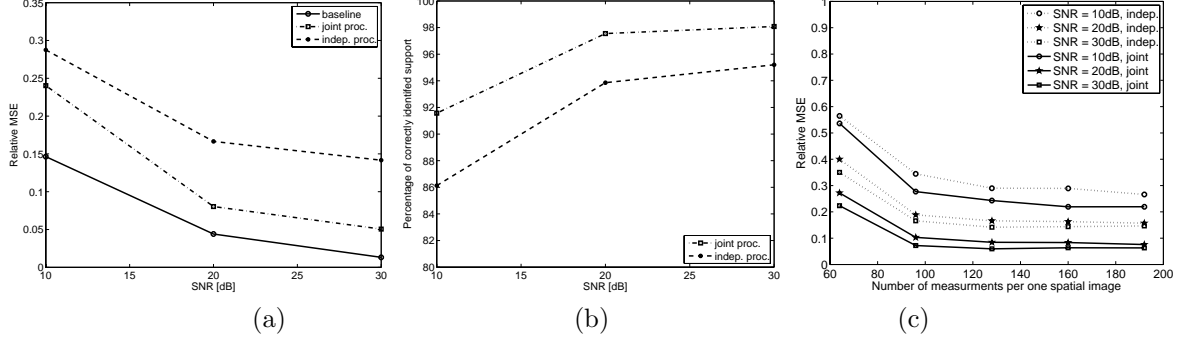

Figure 4. (a) Average relative mean square error as a function of signal to noise ratio (SNR) for 20 Monte Carlo runs with 8 phase history aspects per one spatial image, (b) percentage of correctly identified support and (c) average $RMSE$ sensitivity to number of measurements per one 16×16 spatial image.

where we visually compare different pixel errors when averaged across azimuth. As expected, joint processing strongly suppresses errors outside set T, whereas noise level outside the set T is increased for independent point enhanced processing. Typical anisotropic responses and their reconstructions of several pixels over a full range of 20 aspect angles are shown in Figure 6. Independent reconstruction introduces spurious responses at 'non-active' pixel locations and at times it misses to identify certain azimuth responses. Noise floor of pixels in the set T that have zero response at certain azimuths is typically smaller for the independent reconstruction (the top row of Figure 6). On the other hand, noise floor at pixels in the complement set T^c is smaller for joint processing (the last two figures in the bottom row of Figure 6). This is an expected behavior as joint processing explicitly imposes sparsity on pixels in the set T^c.

In Figure 4(c) we show the SAR sensor matrix sensitivity to reduction in the number of observation aspects. Azimuth returns are subsampled uniformly at random, such there is the same number of measurements in each
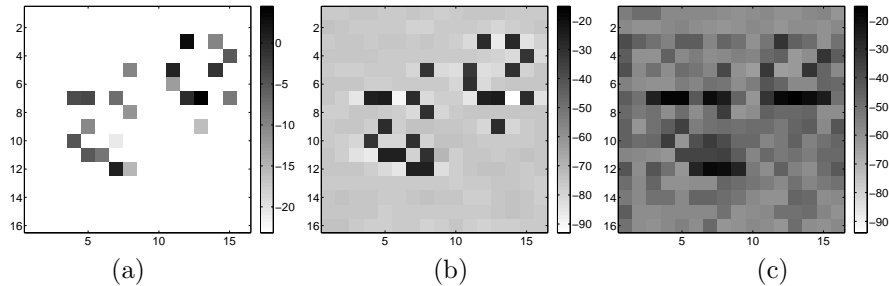

Figure 5. (a) Average spatial magnitude of reflectivity response over azimuth for the sample ground truth, (b) average spatial error over azimuth for joint processing and (c) average spatial error over azimuth for independent processing ($SNR = 20$dB).
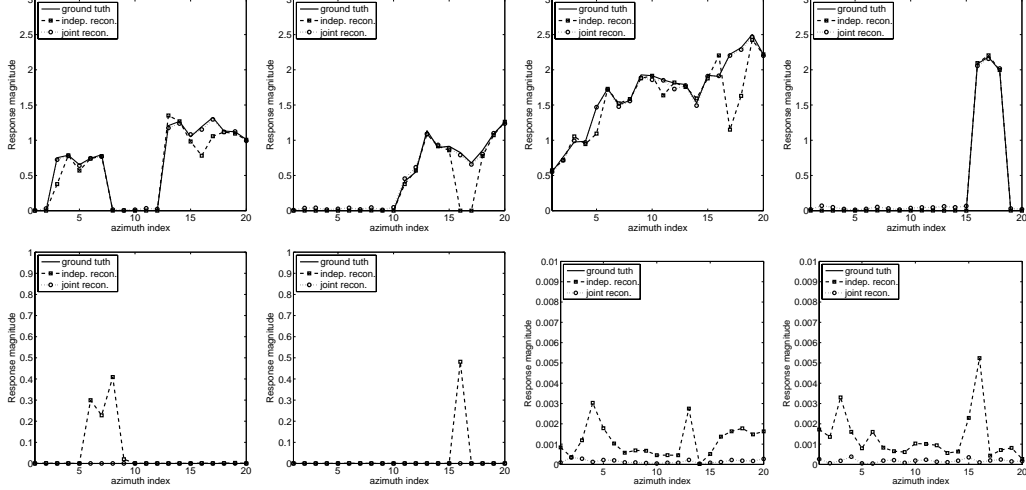
Figure 6. Reflectivity magnitude vs azimuth for several sample pixels in the support set T (top row) and the set T^c (bottom row). Note that pixels in the set T^c have zero true response.

sub-aperture. From a theoretical point of view and under certain assumptions on a linear forward operator, the number of measurements needed to reconstruct a sparse signal is proportional to its support size, rather that its cardinality.[11] The result in Figure 4(c) indicates that SAR forward operator $\boldsymbol{\Phi}$ falls into the category of operators that allow for measurement compression, i.e. there is no cost in the achieved $RMSE$ for the wide range of azimuth sub-sampled measurements.

4.3 Backhoe Xpatch data set reconstruction

In this section we present imaging results based on a backhoe dataset, generated by the XPatch simulator.[9] The CAD model of the backhoe is given on Figure 7(a). The phase history data are collected over $\Delta\theta = 110^o$ azimuths in the range $[-10^o, 100^o]$ at 30° elevation, with the frequency bandwidth $B = 500$MHz around the center frequency $f_c = 10$GHz. The reconstruction grid is chosen such that one 128×128 spatial image is reconstructed every 5°. Thus, there are total of 22 jointly reconstructed images corresponding to 22 consecutive, non-overlapping viewing aspects.

First, we apply the traditional polar-format algorithm on phase history over the full range of aspects and reconstruct one image, Figure 7(b). Polar format algorithm is implemented by applying 1-D range resampling, followed by 1-D azimuth resampling.[4] In order to avoid ringing in the spatial domain due to limited bandwidth in spatial frequency (wavenumber) domain, we apply Taylor windowing on the resampled data before taking the 2-D inverse Fourier transform. Taylor window is specified by 4 nearly constant-level sidelobes adjacent to the mainlobe and -35dB sidelobe suppression below the mainlobe level.

Due to visualization constraints we first present a composite WSAR image obtained by independent-point enhanced processing in Figure 7(c) and the composite image obtained by joint processing in Figure 7(d). The composite image is defined as an image of maximum pixel reflectivity magnitudes across all azimuths.[2] This simple metric aims at finding the peak response across all viewing angles of a spatial pixel (x_n, y_n), i.e. $\max\{|s(x_n, y_n, \theta_i)|, i = 1, \ldots, I\}^2$. Note that these images are plotted in dB scale, by first thresholding small values to zero at the same threshold level for both joint and independent reconstructions. The composite image results show the backhoe's reflectivity in much finer detail when compared to results of polar format algorithm applied to the full aperture data. Spatial support of the jointly reconstructed composite image is much smaller and only the dominant features are reconstructed. Independent reconstructions also identify dominant features similarly, however some spurious scatterers appear to be present in the reconstructed image.

Next, in Figure 8 we present magnitudes of the backhoe's spatial reflectivity when viewed from several consecutive reconstruction angles. Joint and independent, point-enhanced processing produce better focused
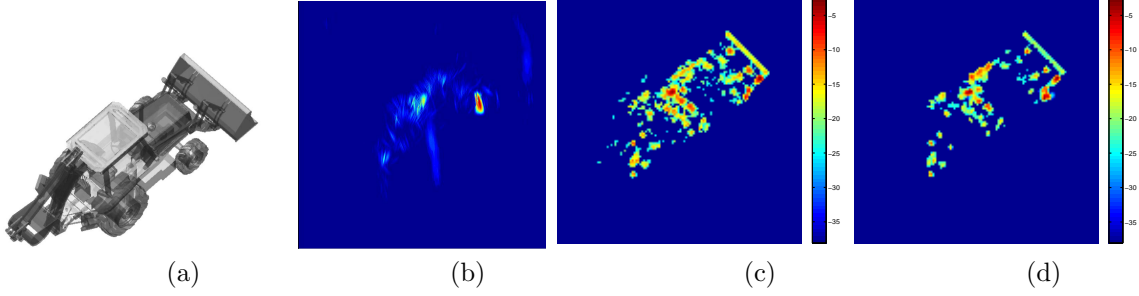
Figure 7. (a) The backhoe CAD model, (b) polar format algorithm applied on the full aperture of 110° and composite images of (c) independent, point-enhanced reconstruction ($q = .8, \beta = .1$) and (d) joint reconstruction ($p = .8, \alpha = .05, q = .8, \beta = .1$) of 22 images each corresponding to a sub-aperture of 5°.
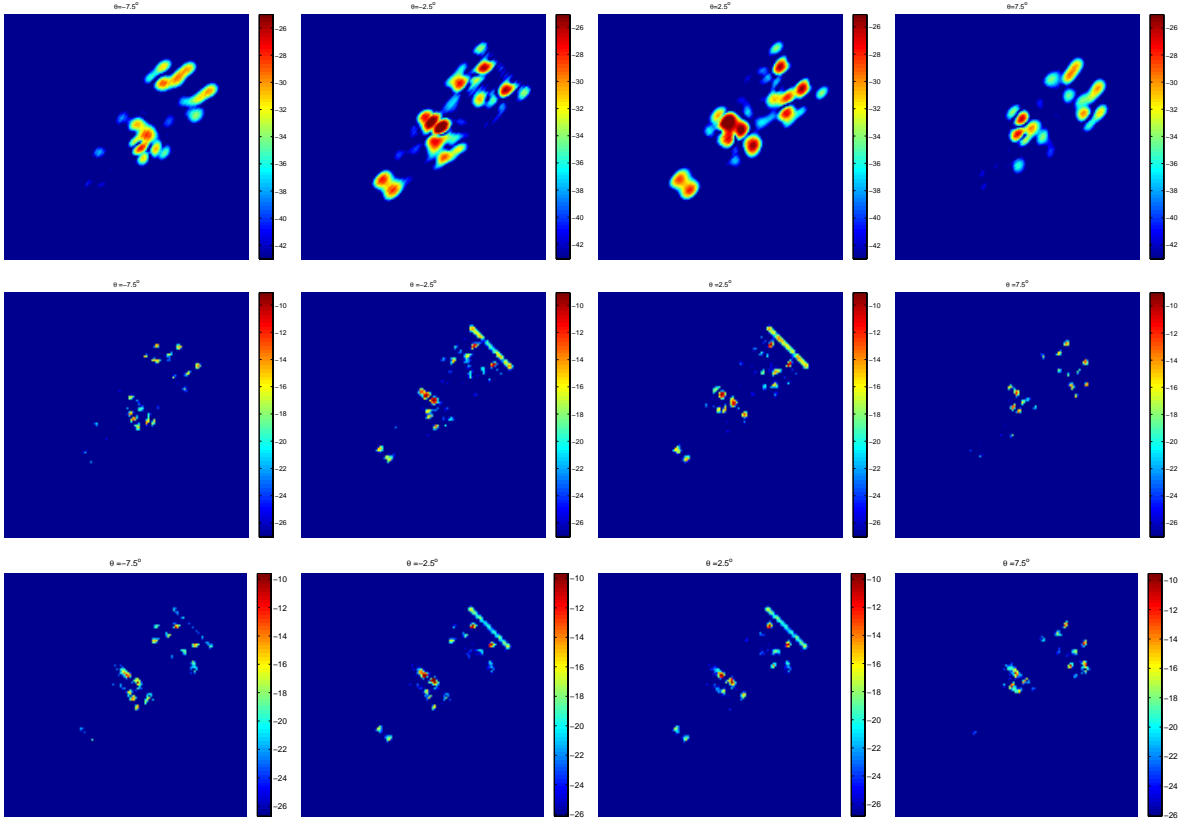

Figure 8. Three sample reconstructed SAR images each of 5° width with maximum number of measurements. Columns left to right correspond to images centered at -7.5^o, -2.5^o, 2.5^o, 7.5^o degrees azimuth. Rows correspond to polar format with Taylor windowing reconstruction (top row), independent (middle row) and joint processing (bottom row).

imagery, whereas a noticeable point spreading is visible at the images reconstructed by the polar format algorithm. Joint and independent reconstructions are plotted on the same dB scale. Independent reconstruction yields larger magnitude responses, i.e. point-enhancing, while joint processing produces images with more compact spatial support. Contrasting independent and joint reconstructions of the first three columns of Figure 8, we see smoother change in reflectivities over angle in the joint processing result.

Figure 9 shows reconstructed reflectivity shapes as a function of azimuth for a set of sample pixels. As expected, reflectivity aspect signature has limited persistence, with high correlation over small aspect extents. The fine detail provided in these plots allows for a scattering center feature extraction. For example, scatterers
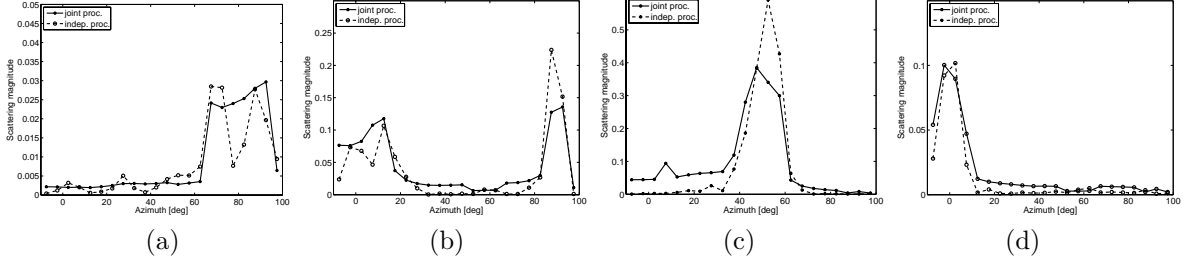
Figure 9. Magnitude of reflectivity response over full range of aspects for several sample pixels.
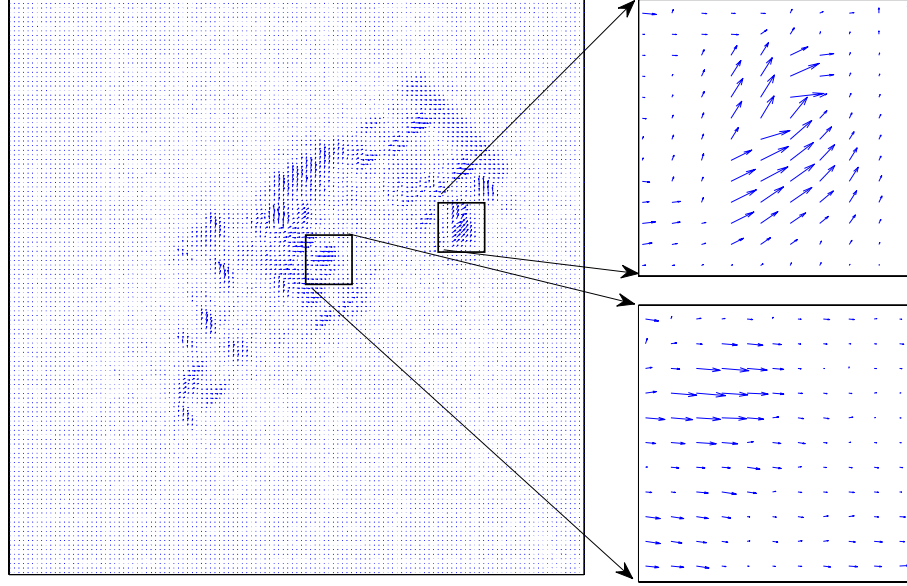

Figure 10. Quiver plot indicating aspect angle of the maximum scattering magnitude response.

such as flat, metal plates have glint anisotropy that is very thin in azimuth, whereas flag and metal poles act as isotropic point scatters. Note that joint processing typically produces smooth scattering shapes, whereas independent processing reconstructs shapes that are jittery. Similarly to the synthetic example, a noise floor in azimuth direction, at point scatterer locations, appears somewhat elevated in the case of joint processing.

In Figure 10 we present a quiver plot indicating aspect angle of the maximum scattering magnitude response.

In Figure 11 we show a set of composite image reconstructions for a sparse collection aperture. In particular, for joint and independent processing the sparse aperture is defined as azimuth subsampled phase history returns. Phase history azimuths within each sub-aperture, i.e. for each image, are chosen uniformly at random among full set of azimuth returns such that each image has equal number of measurements. In contrast, subsampling for the polar format algorithm is performed uniformly, but non-random to aid range and azimuth resampling of the phase history returns. Random downsampling with polar format reconstructions produces much worse results and we omit presenting these plots. We first see that the quality of the composite reflectivity image reconstruction is weakly dependent on the number of azimuth measurements. Joint and independent processing appear more robust when compared to polar format reconstructions. As the subsampling drops down to 35%, independent, point-enhanced processing tends to increase a number of spurious point scatterers , whereas spatial support for joint processing remains focused with further point sharpening.
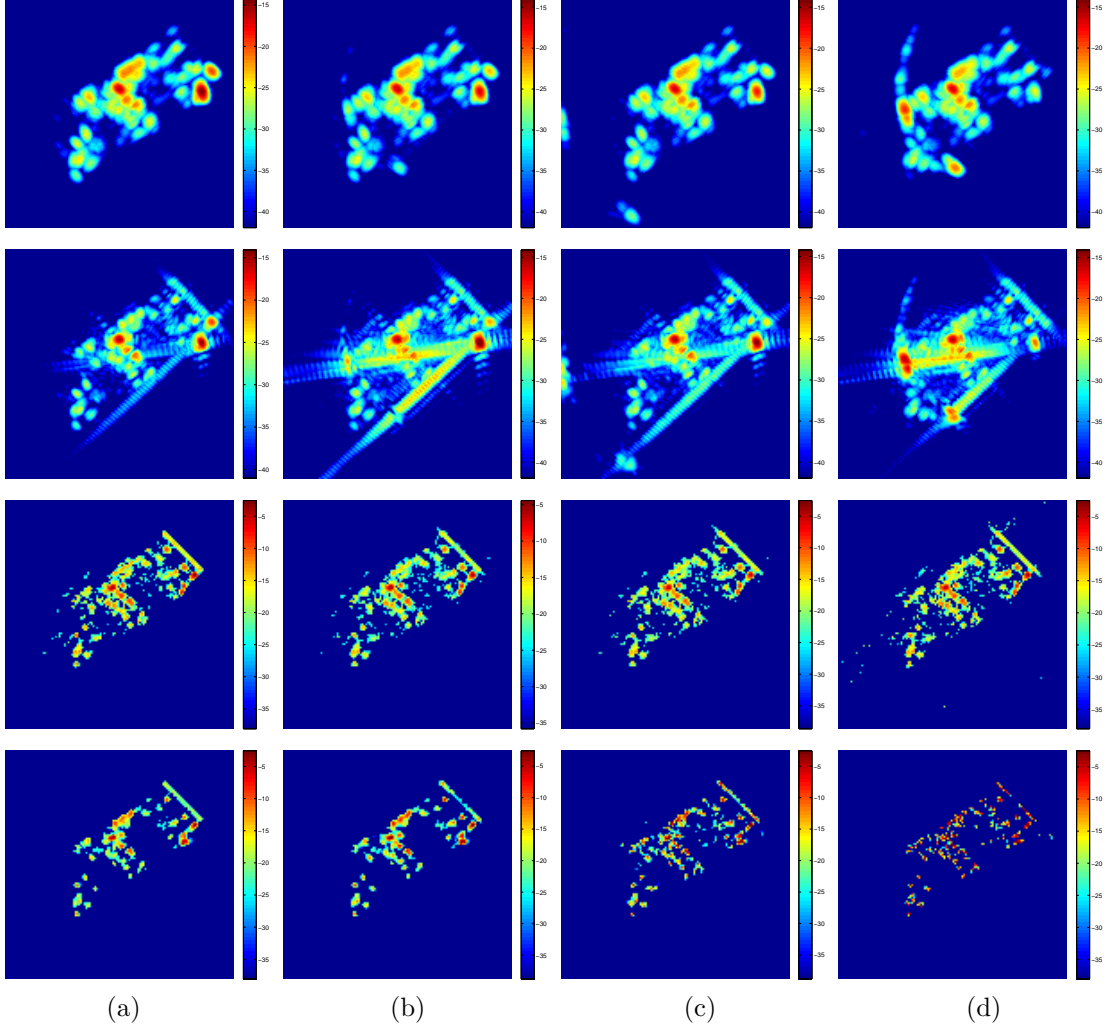
Figure 11. Composite WSAR images with azimuth phase history returns sub-sampled at (a) 100%, (b) 70%, (c) 50% and (d) 35% of the maximum number of available azimuth measurements. Composite images correspond to polar format algorithm with and without Taylor windowing(top two rows), independent, point enhanced processing (third row) and joint processing (bottom row).

5. CONCLUSION

We have approached wide-angle SAR reflectivity reconstruction as a three-dimensional inverse problem exploiting the fact that spatial reflectivity fields are sparse and that their magnitudes are smooth with fast transitions at random aspect angles. This approach allows for anisotropic reflectivity characterization without the need for detailed prior knowledge of azimuth persistence or scattering type. We have shown that this algorithm produces better focused imagery on Xpatch Backhoe data set when compared to traditional polar format algorithm. Algorithms that can finely characterize anisotropy of the scene's reflectivity field, provide a path for moving from pixel-based imaging to object level information extraction. This information can be tied to higher processing blocks that perform e.g. automatic target recognition (ATR). Furthermore, reconstruction quality exhibits robustness to limitations in data quantity, leaving room for a spotlight SAR sensor to multiplex interrogation of more than one ground scene during phase history collection.

APPENDIX A. ALGORITHM

The solution to the minimization problem can be obtain multiple ways and here we present an algorithm based on a quasi-Newton method. For general $0 \leq p \leq 1$, the l_p norm is non-differentiable around the origin. In the first step, any l_p norm is approximated by a smooth function $\|z\|_p^p \approx \sum_{i=1}^{K} ((z_i)^2 + \epsilon)^{\frac{p}{2}}$. The gradient can be written in compact form:

$$\Delta J_\epsilon(\mathbf{s}) = H(\mathbf{s})\mathbf{s} - 2\mathbf{\Phi}^H \mathbf{r} \tag{8}$$

where Hessian approximation $H(\mathbf{s})$ is given by:

$$H(\mathbf{s}) = 2\mathbf{\Phi}^H\mathbf{\Phi} + p\alpha P^H(\mathbf{s})\mathbf{D}_\theta^T \Lambda_1(\mathbf{s})\mathbf{D}_\theta P(\mathbf{s}) + q\beta\Lambda(\mathbf{s})$$
$$\Lambda_1(\mathbf{s}) = \text{diag}\{\|(\mathbf{D}_\theta|\mathbf{s}|)_k\|^2 + \epsilon)^{p/2-1}\}$$
$$\mathbf{D}_\theta = \text{diag}\{[-I, I]\}$$
$$P(\mathbf{s}) = \text{diag}\{\exp(-j\angle(\mathbf{s})_k)\}$$
$$\Lambda(\mathbf{s}) = \text{diag}\{\text{diag}\{(\sum_{i=1}^{I} |s(x_n, y_n, \theta_i|^2 + \epsilon)^{q/2-1}\})\}$$

The quasi-Newton solution at iteration m is

$$\widehat{\mathbf{s}}^{(m+1)} = \widehat{\mathbf{s}}^{(m)} - \delta[H(\widehat{\mathbf{s}}^{(m)})]^{-1}\Delta J_\epsilon(\widehat{\mathbf{s}}^{(m)}), \tag{9}$$

where δ controls a size of the quasi-Newton step. Substituting the gradient of the cost function given in Equation 8, the quasi-Newton iteration is given by

$$H(\widehat{\mathbf{s}}^{(m)})\widehat{\mathbf{s}}^{(m+1)} = (1-\delta)H(\widehat{\mathbf{s}}^{(m)})\widehat{\mathbf{s}}^{(m)} + \delta 2\mathbf{\Phi}^H\mathbf{r}. \tag{10}$$

Note that this a linear set of equations with the unknown $\widehat{\mathbf{s}}^{(m+1)}$ and the right hand side recalculated at each iteration. This system can be solved itself iteratively by for example, a conjugate gradient method.

REFERENCES

[1] Potter, L. C. and Moses, R. L., "Centers for SAR ATR," *IEEE Transactions on Image Processing* (January 1997).

[2] Moses, R. L., Potter, L. C., and Cetin, M., "Wide angle SAR imaging," *SPIE Defense and Security Symposium, Algorithms for Synthetic Aperture Radar Imagery XI* (April 2004).

[3] Varshney, K. R., Cetin, M., FisherIII, J. W., and Willsky, A. S., "Joint image formation and anisotropy characterization in wide-angle SAR," *SPIE Defense and Security Symposium, Algorithms for Synthetic Aperture Radar Imagery XIII* (April 2006).

[4] C.V.Jakowatz, E.Wahl, D., S.Eichel, P., C.Ghiglia, D., and Thompson, P. A., [*Spotlight-mode Synthetic Aperture Radar: a Signal Processing Approach*], Kluwer Academic Publishers, Norwell, MA (1996).

[5] Bertero, M., "Linear inverse and ill-posed problems," *Advances in Electronics and Electron Physics* **75**, 1–120 (1989).

[6] Malioutov, D. M., Cetin, M., and Willsky, A. S., "A sparse signal reconstruction perspective for source localization with sensor array," *IEEE Trans. Signal Processing* **53**, 3010–3022 (August 2005).

[7] Cotter, S. F., Rao, B. D., Kjersti, E., and Kreutz-Delgado, K., "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing* **53**, 2477 – 2488 (July 2005).

[8] Cetin, M., [*Feature-Enhanced Synthetic Aperture Radar Imaging*], Ph.D. Thesis, Boston University (February 2001).

[9] "Backhoe data sample and visual-d challenge problem," *Air Force Research Laboratory, https://www.sdms.afrl.af.mil/main.htm* .

[10] Donoho, D. L., Johnstone, I. M., Stern, A. S., and Hoch, J. C., "Maximum entropy and the nearly black object," *Journal of the Royal Statistical Society B* **54**, 41–81 (1992).

[11] Candes, E. J., "Compressive sampling," *Int. Congress of Mathematics* **3**, 1433–1452 (2006).